CC   expression of the polypeptide. Antagonists, inhibitory nucleic acid or
CC   competitive polypeptide are useful for inhibiting the polypeptide e.g.
CC   bacterial (especially S. aureus) infections. They are also useful against
CC   Helicobacter pylori infections and related cancers, ulcers and gastritis.
CC   The antibacterial agents are useful to treat in-dwelling devices for
CC   infection prevention or generally as wound treatments to prevent adhesion
CC   of bacteria to matrix proteins. The MurC polypeptide is also useful for
CC   diagnosing or prognosing a (susceptibility to) disease, for raising
CC   antibodies; to identify modulators or specific receptors; in rational
CC   drug design and as an immunogen for vaccines. The MurC gene sequences are
CC   useful in antisense/ribozyme therapeutics; to detect mutant MurC gene;
CC   for chromosomal mapping; to determine bacterial serotype; and for genetic
CC   immunisation.
SQ   Sequence 1351 BP; 471 A; 166 C; 260 G; 454 T;

Query Match   100.0%; Score 1351; DB 60; Length 1351;
Best Local Similarity 100.0%; Pred. No. 0.00e+00;
Matches 1351; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Db   1 atgagtaaggagtttatataatgacacactatcatttgtcggaattaaaggttctggc 60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1 ATGAGTAAGGAGTTTATATAATGACACACTATCATTTGTCGGAATTAAAGGTTCTGGC 60

Db   61 atgagttcattagcacaaatcatgcatgattaggacatgaagttcaaggatcggatatt 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   61 ATGAGTTCATTAGCACAAATCATGCATGATTAGGACATGAAGTTCAAGGATCGGATATT 120

Db   121 gagaactacgtatttacagaagttgctcttagaaataaggggataaaatattaccattt 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   121 GAGAACTACGTATTTACAGAAGTTGCTCTTAGAAATAAGGGGATAAAATATTACCATTT 180

Db   181 ggtgctaataacataaaagaagatatggtagttatacaaggtaatgcattcgcgagtagc 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   181 GGTGCTAATAACATAAAAGAAGATATGGTAGTTATACAAGGTAATGCATTCGCGAGTAGC 240

Db   241 catgaagaaatagtacgtgcacatcaattgagatgttgtaagttatataatgatttt 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   241 CATGAAGAAATAGTACGTGCACATCAATTGAAATTAGATGTTGTAAGTTATAATGATTTT 300

Db   301 ttaggacagattattgatcaatatacttcagtagcgtaactggtgcacatggtaaaact 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   301 TTAGGACAGATTATTGATCAATATACTTCAGTAGCGTAACTGGTGCACATGGTAAAACT 360

Db   361 tctacaacagtttattatcacacatgttatgaatggtgataaaaaggacttcatttttaatt 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   361 TCTACAACAGTTTATTATCACACATGTTATGAATGGTGATAAAAAGGACTTCATTTTTAATT 420

Db   421 ggtgatggcacaggtgatggatgtgctgaaagtgattatttcgcttttgaggcatgtgaa 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   421 GGTGATGGCACAGGTGATGGGATTGCCTGAAAGTGATTATTTCGCTTTTGAGGCATGTGAA 480

Db   481 tatacgtcactttttaagttataaacctgattacgcaattatgacaaatattgatttc 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   481 TATACGTCACTTTTTAAGTTATAAACCTGATTACGCAATTATGACAAATATTGATTTC 540

Db   541 gatcatcctgattattcaaagatattaatgatgtttgatgcattccaagaaatggca 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   541 GATCATCCTGATTATTCAAAGATATTAATGATGTTTGATGCATTCCAAGAAATGGCA 600

Db   601 cataatgttaaaaaggtatttactactatcgatttaaagattcggatgacattatgtat 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   601 CATAATGTTAAAAAGGTATTTACTACTATCGGATTTAAAGATTCGGATGACATTATGTAT 660

Db   661 gcagatgttccaatttattactatatggatttatagaagattcggatgacattatgat 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   661 GCAGATGTTCCAATTTATTATTACTATATGGATTTATAGAAGATTCGGATGACATTATGAT 720

Db   721 attcaaattacggataaagtactgcttttgatgtgtgatggtggtgagttttatgat 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   721 ATTCAAATTACGGATAAAGGTACTGCTTTTGATGTGTGATGGTGGTGAGTTTTATGAT 780

Db   781 cacttcctgcgtcctccacaatggtgaccacagttttaaatgcattagctgtaattgcg 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy   781 CACTTCCTGCGTCCTCCACAATGGTGACCACAGTTTAAATGCATTAGCTGTAATTGCG 840
Db   841 attagttattagagaagctagagtgttacaaatattaagaagcattagaaacgtttggt 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   841 ATTAGTTATTAGAGAAGCTAGAGTGTTACAAATATTAAGAAGCATTAGAAACGTTTGGT 900

Db   901 ggtgttaaacgtcgtttcaatgaaactacaattgcaaatacaagttattgtagatgattat 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   901 GGTGTTAAACGTCGTTCAATGAAACTACAATTGCAAATCAAGTTATTGTAGATGATTAT 960

Db   961 gcacaccatcaagagaaatagtgctacaattgaccacgcacaagaaatatccacat 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   961 GCACACCATCAAGAGAAATATCAAGAGAAATTAGTGCTACAATTGACCACGCACGGAAGAAATATCCACAT 1020

Db   1021 aaagaagttgttgcagtatttcaaccaccacacttcctagaacaccaagcattttaaat 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1021 AAAGAAGTTGTTGCAGTATTTCAACCACCACACTTCTCAGAACAACAAGCATTTTTAAAT 1080

Db   1081 gaattgcagaagtttatgtaaagcagatcgtgtattctatgtgaaattttggctca 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1081 GAATTTGCAGAAGTTTATGTAAAGCAGATCGTGTATTCTATGTGAAATTTTTGGCTCA 1140

Db   1141 attagagaaaattcggcgccattacgatacaagatttaattgataaaattggaagtgca 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1141 ATTAGAGAAAATTCTGGCGCCATTAACGATACAAGATTTAATTGATAAAATTGGAGGTGCA 1200

Db   1201 tcgttcattaatgaagatctttattagtgattagaacaatttgatcatggtgtttta 1260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1201 TCGTTCATTAATGAAGATCTTATTAGTGATTAGAACAATTTGATCATGGTGTTTTTA 1260

Db   1261 tttatgggtgcaggtgtatattcaaaaattacaaaatgcatatttgataatattaggcatg 1320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1261 TTTATGGGTGCAGGTGTATATTCAAAAATTACAAAAATGCGTATTTTAGATAAAATTAGGCATG 1320

Db   1321 aaaatgcgtttttatatgttttataatagag 1351
            ||||||||||||||||||||||||||||||
Qy   1321 AAAAATGCGTTTTAAATATGTTTATAATAGAG 1351

```
Qy  781 CACTTCCTGTCTCCACAATATGTGACCATACAGTTTTAAATGCATTAGCTGTAATTGCG  840
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1446 attagttattagaagctagagtgttacaaatattaagaagcattagaaacgtttggt  1505

Qy  841 ATTAGTTATTAGAGAAGCTAGATGTTACAAAATATTAAGAAGCATTAGAAACGTTTGGT  900
        |||||||||||||||||||||||||||||||||||||||||||||||||
Db 1506 ggtgttaaacgtcgttcaatgaaactacaatcaagttattgtagatgattat  1565

Qy  901 GGTGTTAAACGTCGTTCAATGAAATCAATGCAAAATCACAATCAAGTATTCGAGATGATTAT  960
        |||||||||||||||||||||||||||||||||||||||||||||||||
Db 1566 gcacaccatccaagaagaaattagtgctacaattgaacagcacgacgaaagaaatatccac  1625

Qy  961 GCACACCATCCAAGAGAATTAGTGCTACAATTGACACGAATTGACAAAGAAAATAATCCAT  1020
        ||||||| ||||||||||||||||||||||||||||||||||||
Db 1626 aaagaagtgttgcagtattctcaccacacacttctctagacacaggcattttaat  1685

Qy 1021 AAAGAAGTGTTGCAGTATTTCAACCACCACACTTCTCTAGAACCACAAGCATTTTTAAAT  1080

Db 1686 gaatttgcagaagttaagtaaagcagatcgtgtattcttatgtgaattcttatgtggatca  1745
        |||||||||||| ||||| |||||||||||||||| ||||||| || |||
Qy 1081 GAATTTGCAGAAAGTTTATGTAAAGCAGATCGTGTATTCTTATGTGAAATTTTTGGCTCA  1140

Db 1746 attagagaaaatactggcgcattaacgatacaagattaattgataaaatgaaggtgca  1805
        ||||||||||| |||||||||||||||||||||| |||||||||| |||| |||||||
Qy 1141 ATTAGAGAAAAATTCTGGCGCATTAACGATACAAGATTTAATTGATAAAATTGGAGGTGCA  1200

Db 1806 tcgttaattaatgaagattctattaatgtattagaacaattgataatgctgttattta  1865
        ||||| |||||||||||| ||||||||| || ||||||||||||||||||||||| ||||
Qy 1201 TCGTTCAATTAATGAAGATCTTATTAATGTTAATGATTAGAAGAACAATTTGATAATGCTGTGTTTTA  1260

Db 1866 tttatgggtgcagttgattgattattcaaaaattacaaaatgcatattagataaaattaggcatg  1925
        |||||||||||||||||| ||||||||||||| |||||||| |||||| ||||||||||||||
Qy 1261 TTTATGGGTGCAGTTGATTATTCAAAAATTACAAAAATGCATATTTAGATAAAATTAGGCATG  1320

Db 1926 aaaaatgcgtttaatatgtttaatatagag  1956
        |||||||||||| |||||||||||| |||||
Qy 1321 AAAAATGCGTTTTAATATGTTTAATAAATAGAG  1351
```

RESULT 3
ID   V80065 standard; DNA; 660 BP.
AC   V80065;
DT   17-MAR-1999 (first entry)
DE   Partial nucleotide sequence of the MurC gene.
KW   MurC gene; UDP-N-acetylmuramate:L-alanine ligase; MurC polypeptide;
KW   bacterial; infection; H. pylori; cancer; ulcer; gastritis; vaccine;
KW   immunogen; drug; genetic immunisation; ds.
OS   Staphylococcus aureus.
FH   Key   Location/Qualifiers
FT   CDS   2..660
FT   /tag= a
FT   /product= "partial MurC polypeptide"
PN   EP-889123-A2.
PD   07-JAN-1999.
PF   26-JUN-1998; US-052720.
PR   03-JUL-1997; US-052720.
PA   (SMIK) SMITHKLINE BEECHAM CORP.
PA   (SMIK) SMITHKLINE BEECHAM PLC.
PI   Burnham MKR, Wallis NG;
DR   WPI; 99-062655/06.
DR   P-PSDB; W89199.
PT   New isolated MurC polypeptide from Staphylococcus aureus and related
PT   nucleic acid - useful in diagnosis, treatment and prevention of
PT   bacterial infections
PS   Claim 2; Pages 4-5; 39pp; English.
CC   The invention relates to a UDP-N-acetylmuramate:L-alanine ligase
CC   (MurC polypeptide) encoded by the S. aureus MurC gene. Host cells
CC   containing an expression system comprising the MurC gene can be used for
CC   the recombinant production of the polypeptide. Agonists or the MurC
CC   polypeptide are used to treat conditions requiring increased activity or
CC   expression of the polypeptide. Antagonists, inhibitory nucleic acid or
CC   competitive polypeptide are useful for inhibiting the polypeptide e.g.
CC   bacterial (especially S. aureus) infections. They are also useful against

CC   that protein-encoding or regulatory regions of commercial, therapeutic or
CC   industrial importance can be obtained Specifically, sequences which are
CC   likely to encode antigens have been identified and these polypeptides can
CC   be used in a vaccine composition against s.aureus infection. The
CC   polypeptides can also be used in a kit for the immunodetection of
CC   S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC   including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC   skin and surgical wound infections, scalded skin syndrome, toxic shock
CC   syndrome, etc. Organisms transformed with the DNA sequences can be used
CC   for recombinant production of the polypeptides. The new DNA sequences
CC   (and their fragments) are useful as primers or probes for isolating
CC   homologues of any of the s.aureus DNA sequences contained on the
CC   computer readable medium.
SQ   Sequence 2424 BP; 839 A; 302 C; 462 G; 759 T;

Query Match   93.8%; Score 1267; DB 58; Length 2424;
Best Local Similarity 94.7%; Pred. No. 0.00e+00;
Matches 1279; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

```
Db  606 atgatgaggaggttttataatgacacactcatttgtcggaattaaagttctggc  665
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1 ATGAGTAAGGAGGTTTTATATAATGACACACTATCATTTGTCGGAATTAAAGGTTCTGGC  60

Db  666 atagttcattagcacaaatcatgcatgttaggacgtgaagttcaaggatcggatatt  725
         |||||||||||||| |||||||||||||||||||||||||||||||||||||||
Qy   61 ATGAGTTCATTAGCACAAATCATGCATGAAGTTCAAGGATCGGAATT  120

Db  726 gagaactacgtattacagaagttgcctctagaaataagggataaaaatattaccattt  785
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  121 GAGAACTACGTATTACAGAAGTTGCTCTTAGAAATAAGGGATAAAAATATTACCATTT  180

Db  786 gatgctaataacataaagaagatatgtagttatacaggtaatgaatgcgtcggagtagc  845
         | |||||||||||||||||||||||||||||||||||||||||||||||
Qy  181 GGTGCTAATAACATAAAGAAGATATGTAGTGTAGTTATACCAGGTAATGCCATTCGCGAGTAGC  240

Db  846 catgaagaagtagtacgtgcacactcaattgaaattgatagtgttgtaagttataatgattt  905
         |||||||| |||||| |||||||| || ||||| ||| || |||| |||||||||||||
Qy  241 CATGAAGAAATAGTAGTACGTGCACATCATCGCACATCAATTGAAATTAGATGTTGTAAGTTATAATGATTTT  300

Db  906 taggacagatattgatcaaatatacctcagtagctgtaactggtgcacatgtaaaact  965
         |||||||| ||||||||| |||| |||||||||||||||||||||||| ||||||
Qy  301 TTAGGACAGATTATTGATCAATTACTTCAGTAGCTGTAACTGGTGCACATGTGTAAAACT  360

Db  966 tctacaacaggttattatcacatgttatgatggtgataaaaagacttcattttattt  1025
         ||||||||||||||| |||||||| ||| ||||| |||||||||||||||||| ||||
Qy  361 TCTACAACAGGTTTATTATCACATGTTTATTCACATGTATGGTGATAAAAAGACTTCATTTTTAATT  420

Db 1026 ggtgatggcacaggtatggattggtctgaaagtgatgattattcgctttgaggcatgtaa  1085
         |||||||||||||||||| ||| ||||||||| |||| ||||||||| ||| |||||
Qy  421 GGTGATGGCACAGGTATGGGATTGGCCTGAAAGTGATTATTTCGCTTTTGAGGCATGTGAA  480

Db 1086 tatagagctcactttttaagttataaacctgattacgcaattatgacaaaatattgnnnnn  1145
         |||||||||||||||||||||||||||||| |||||| ||| |||||| |||| |||
Qy  481 TATAGAGCTCACTTTTTAAGTTATAAAACCTGATTATAAACCTGACAAATATTGATTTC  540

Db 1146 nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnntggca  1205
         |||
Qy  541 GATCATCCTGATTATTCAAAGATATTAATGATGTTTTGATGCATTCCAAGAAATGGCA  600

Db 1206 cataatgttaaaaaaggtattattgcttgggcgatgatgaacatcacgtaaaattgaa  1265
         |||||||||||||||| ||| |||||||||| |||||||||||| ||||||||||||
Qy  601 CATAATGTTAAAAAAGGTAATATTGCTTGGGGTGATGATGAACATCTACGTAAAATTGAA  660

Db 1266 gcagatgttccaattattattatgatggattaaagatttatgtgctcaaat  1325
         |||||||||||||||||| | |||||||| ||||||| ||||| |||||||
Qy  661 GCAGATGTTCCAATTATTACTAATGGATTTAAAGATATTCGGATGACATTTATCTCAAAT  720

Db 1326 attcaaattacggataaaggtactgctttgatgtgtatggtggtgagtttatgat  1385
         |||||| |||||||||||||||||||||| |||||||||||||| ||||| ||||
Qy  721 ATTCAAAATTACGGATAAAGGTACTGCTTTTGATGTGTATGGTGGAGTTTTATGAT  780

Db 1386 cacttcctgtctccacaatgtgaccatacagtttaaatgcattagctgtaattgcg  1445
         |||||||||||||||||||||||| |||||||||| ||||||||||||||||||||
Qy  781 (continued above)
```

```
CC    expression of the polypeptide. Antagonists, inhibitory nucleic acid or
CC    competitive polypeptide are useful for inhibiting the polypeptide e.g.
CC    bacterial (especially S. aureus) infections. They are also useful against
CC    Helicobacter Pylori infections and related cancers, ulcers and gastritis.
CC    The antibacterial agents are useful to treat in-dwelling devices for
CC    infection prevention or generally as wound treatments to prevent adhesion
CC    of bacteria to matrix proteins. The MurC polypeptide is also useful for
CC    diagnosing or prognosing a (susceptibility to) disease, for raising
CC    antibodies; to identify modulators or specific receptors; in rational
CC    drug design and as an immunogen for vaccines. The MurC gene sequences are
CC    useful in antisense/ribozyme therapeutics; to detect mutant MurC gene;
CC    for chromosomal mapping; to determine bacterial serotype; and for genetic
CC    immunisation.
SQ    Sequence   1351 BP;   471 A;   166 C;   260 G;   454 T;

Query Match      98.4%;   Score 5363; DB 60;   Length 1351;
Best Local Similarity   60.9%;   Pred. No. 0.00e+00;
Matches   799;  Conservative   320;  Mismatches 192;  Indels   0;  Gaps   0;
```

```
Db      22  atgacacactaccatttgtcggaattaaaggttctgcatgagtccattagcacaaatc    81
             ||:|| ||::|| ||| |||:||| ::|||||||| ||  :: :| :|||:| :::::||
Qy       1  ATGACNCAYTAYCAYTTYGTNGGNATHAARGGNWSNGGNATGWSNWSNYTNGCNCARATH    60

Db      82  atgcatgattaggacatgaagttcaaggatcggatattgagaactacgtatttacagaa   141
             |||||||:|| ||:|| ||  || |||||| ||:|| ::|||:|||||| ||||:||
Qy      61  ATGCAYGAYYTNGGNCAYGARGTNCARGGNWSNGAYATHGARAAYTAYGTNTTYACNGAR   120

Db     142  gttgtcctgagaaataagggcgtaaaaatattaccatttgtgctaataacataaaagaa   201
             || |  :: ||| ||| ||:|| ||| ||||:|| || || |||||::| :||||||||
Qy     121  GTNGCNYTNMGNAAYAARGGNATHAARATHYTNCCNTTYGGNCNAAYAAYATHAAARGAR   180

Db     202  gatatgggtagttatacaagttatgcattccgccgagtagccatgaagaaatagtacgtgca   261
             || ||||| || ||||| |||||| || ||:|| |||||:|| |||:|| |||||:||
Qy     181  GAYATGGGTNATHCARGGNAAYGCNTTYCNCNWSNWSNCAYGARGARATHGTNMGNGCN   240

Db     262  catcaattgaaattagatgtgtaagttataatgatttttaggacagattattgatcaa   321
             ||:||| ||::|| ||:|| || ::|||||:|| :| |||::||||| ||||||:||
Qy     241  CAYCARYTNAARYTNGAYGTNGTNWSNTAYAAYGAYTTYNGGNCARATHATHGAYCAR   300

Db     322  tatacttcagtagctgtaactggtgcacatggtgcaaacttctacacacaggttattatca   381
             || | :: || || || || || |||||| |||| || :|||:| :|:| |||:|| |
Qy     301  TAYACNWSNGTNGCNGTNACNGGNGCNCAYGGNAARACNWSNACNACNGGNTYNYTNWSN   360

Db     382  catgttatgaatggtgataaaagacctcattttaattggtatgtgccacaggtatggga   441
             |||||:|| ||:||||| | |:|| || ||| |||||| ||:|| |||||||| ||:||
Qy     361  CAYGTNATGAAYGGNGAYAARAARACNWSNTTYTNATHGGNGAYGGNACNGGNATGGGN   420

Db     442  ttgcctgaaagtgattatttcgctttgaggcatgtgaatatagacgtcacttttaagt   501
             :|:||| |||| ||||| ||:|| :||:|| ||:|| ||||| |||:| ::|| ::
Qy     421  YTNCCNGARWSNGAYTAYTTYGCNTTYGARGCNTGYGARATAYMGNMGNCAYTTYTNWSN   480

Db     502  tataaacctgattacgcaattatgacaaatatgattcgatcatccgattattccaa   561
             |||:|| ||:|| || | ||||| |||||||| |||:| |||:|||:|| |||:||||
Qy     481  TAYAARCCNGAYTAYGCNATHATGACNAAYATHGAYTTYGAYCAYCCNGAYTAYTTYAAR   540

Db     562  gatattaatgatgtttgatgcattccaagaaatggccacataatgttaaaaaaggtatt   621
             ||||| || ||:||||| ||:|| || ||||:|| |||:|| |||:|||||| ||||||
Qy     541  GAYATHAAYGAYGTNTGAYGCNTTYCARGARATGGCNCAYCAYAAYGTNAARAARAGGNATH   600

Db     622  atggctggggtgatgataggaacatctacgtaaaattgaagcagagtgttccaattattac   681
             ||:|| ||||||:||||| |||| ||| |||| |||:||| | |||||:|| :|||||||||
Qy     601  ATHGCNTGGGGNGAYGAYCAYYTNMGNAARATHGARGCNGAYGTNCCNATHTAYTAY    660

Db     682  tatggattaaagattcggatgacatttatgctcaaaatattcaaattacggataaaggt   741
             |||||:|| ||:||||| || ||:|| |||||||||:||||:| ::|||||| :|||||
Qy     661  TAYGGNTTYAARGAYWSNGAYGAYATHTAYGCNCARAAYATHCARATHACNGAYAARGGN   720

Db     742  actgctttgatgtgtatggatggtgagttttatgatcacttcctgtctccacaatat   801
             |||||:|| ||| || ||||||| ||:|| ||:|| ||:|| :|:| :|||:|||||||
Qy     721  ACNGCNYTNGAYGTNTAYGGNGAYGGNGARTTYTAYGAYCAYYTYTNWSNCCNCARTAY   780

Db     802  ggtgaccatacagttttaaatgcattagctgtaattgcgattagtgaagttattagaagca   861
             ||:|| ||| ||||:||||| || || ||:|| ||:|| |||:||:|| :|||| |||||
```

```
Qy     781  GGNGAYCAYACNGTNTYNAAYGCNYTNGCNGTNATHGCNATHWSNTAYTNGARARAYTN   840

Db     862  gatgttacaaatattaaagaagcattagaaacgtttggtgttaaagtcgtttcaat     921
             ||:|| ||:|| ||||| ||| || |||||||:|| ||:|| |||||:|| ::||
Qy     841  GAYGTNACNAAYATHAARGARGCNYTNGARACNTTYGGNGGNTNAARMGNMGNTTYAAY   900

Db     922  gaaactacaatgcaatcaagttattgtagatgatatgcacaccatccaagagaaatt   981
             ||::|| ||:|| || ||||| ||:||||||| |||||| |||:|||||||| ||:|
Qy     901  GARACNACNATHGCNAAYCARGTNATHGTNGAYGAYATGCAYACNACNCCNMGNGARATH   960

Db     982  agtgctacaatgtgacacagcacgaaagaaatatcccataaagaagttgttgcagtatt  1041
             ::||||:|| :|||| ||||| |||||:|| :||:| |||||||||| |||:| ||:||
Qy     961  WSNGCNACNATHGAYACNGCNMGNAARAARTAYCCNCAYAARGARGTNGTNGCNGTNTTY  1020

Db    1042  caaccacacactttctctagaaccaaGcaagcattttaaatgaattgcagaaagtttatgt  1101
             || || ||| ::|:|| :| |||:|||:|| |||:| |||| || :||||||| |||:
Qy    1021  CARCCNCAYACNTTYWSNMGNACNCARGGCNTYYTNAAYGARTTYGCNGARWSNYTNTGY  1080

Db    1102  aaagcagatcgtgtattcttatgtgaacatttttggcctcaattagagaaaatcctggcgca  1161
             ||:|| |||||:|| :::|||:|| |||| |||:||  :|| ||| |||| ::||:||:||
Qy    1081  AARGCNGAYMGNGTNTTYTNTGYGARAATHTTYGGNWSNATHMGNGARAAYWSNGGNGCN  1140

Db    1162  ttaacgatacaagattaattgataaaattgaggagtgcatcgtcgttcaataatgaagactt  1221
             :| :|| |||:|| |||:|:|| || |||| ||:|| ||:| :| :|| :|||:|||:|
Qy    1141  YTNACNATHCARGAYYNATHGAYAARATHGGNGGNGCNWSNTTYATHAAYGARGAYYTN  1200

Db    1222  attaatgtattagaacaattgataatgctgtgttttatttatggtgcaggtgatatt  1281
             |||::|| || |||| || |||:| ||:|| |:|| |||:||||| || |||:||
Qy    1201  ATHAAYGTNYTNGARCARTTYGAYAAYGCNGTNGTNYTNTNYTATGGNGCNGGNGAYATH  1260

Db    1282  caaaaattacaaatgcatatttagataaatggcatgaaaatgcgttt  1332
             ||:|||:|| || |||:|::| |||||:|| :||||||||:||:|||
Qy    1261  CARAARYTNCARAAYGCNTAYYTNGAYARYTNGGNATNGAARAAYGCNTTY  1311
```

```
RESULT   2
ID     V74703  standard; DNA; 2424 BP.
AC     V74703;
DT     16-MAR-1999 (first entry)
DE     Staphylococcus aureus contig SEQ ID #392.
KW     Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW     cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW     skin infection; surgical wound infection; scalded skin syndrome;
KW     toxic shock syndrome; ds.
OS     Staphylococcus aureus.
FH     Key          Location/Qualifiers
FT     misc_feature  1141..1200
FT                   /tag= a
FT                   /note= "these bases represent a line of missing text in
FT                           the sequence listing in the specification. They
FT                           are included to maintain the nucleotide numbering
FT                           given in the specification for this DNA sequence"
PN     EP-786519-A2.
PD     30-JUL-1997.
PF     07-JAN-1997; 100117.
PR     05-JAN-1996; US-009861.
PA     (HUMA-) HUMAN GENOME SCI INC.
PI     Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI     Rosen CA;
DR     WPI; 97-374922/35.
PT     Polynucleotide(s) and proteins derived from Staphylococcus aureus
PT     stored on computer readable medium and used in the production of
PT     anti-S.aureus vaccines
PS     Claim 1; Page 1287-1288; 3271pp; English.
CC     This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC     of the invention. The DNA sequences are recorded on a computer readable
CC     medium, preferably selected from a floppy or hard disk, random access
CC     memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC     the S.aureus DNA sequences allows putative functions to be assigned so
CC     that protein-encoding or regulatory regions of commercial, therapeutic or
CC     industrial importance can be obtained. Specifically, sequences which are
CC     likely to encode antigens have been identified and these polypeptides can
CC     be used in a vaccine composition against S.aureus infection. The
```

```
CC    polypeptides can also be used in a kit for the immunodetection of
CC    S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC    including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC    skin and surgical wound infections, scalded skin syndrome, toxic shock
CC    syndrome, etc. Organisms transformed with the DNA sequences can be used
CC    for recombinant production of the polypeptides. The new DNA sequences
CC    (and their fragments) are useful as primers or probes for isolating
CC    homologues of any of the S.aureus DNA sequences contained on the
CC    computer readable medium.
SQ    Sequence   2424 BP;   839 A;   302 C;   462 G;   759 T;

Query Match            91.9%;   Score 5012;   DB 58;   Length 2424;
Best Local Similarity  57.7%;   Pred. No. 0.00e+00;
Matches   757;   Conservative  301;   Mismatches 253;   Indels   0;   Gaps   0;

Db   627 atgacacactcatttgtcggaattaaggttctggcatggagttcattagcacaaatc 686
         ||:||:|:|||:||:||||:||||| |||::||:||:| :::|| ||:||::||||||
Qy     1 ATGACNCAYTAYCAYTTYGTNGGNATHAARGGNWSNGGNATGWSNWSNYTNGCNCARATH 60

Db   687 atgcatgattaggacatgaagttcaaggatcggatattgagaactacgtatttacagaa 746
         ||:||::|||  || :||||||| ::: ||:||||| |||||| || || |||||||||
Qy    61 ATGCAYGAYYTNGGNCAYGARGTNCARGGNWSNGAYATHGARAAYTAYGTNTTYACNGAR 120

Db   747 gttgctcttagaaataagggataaaaattaccatttgatgctaataacataaaagaa 806
         ||:|| ::| | |||||:|| :|||||||||:||:||:||| ||||| ||| |||||
Qy   121 GTNGCNYTNMGNAAYAARGGNATHAARATHYTNCCNTTYGNGCNAAYAAYATHAARGAR 180

Db   807 gatatggtagttacacaggtaatgcattcgcgtagccatgaagaaatagtacgtgca 866
         || |||||| || |||||||:|| || ||:||||:|| |:| || |||||:||:||
Qy   181 GAYATGGTNGTNACHCARGGNAAYGCNTTYGCNWSNSWSCAYGARGARATHGTNMGNGCN 240

Db   867 catcaattgaaattagatgttgtaagttataatgatcttttaggacagattattgatcaa 926
         ||:|:|| ||:| |||| ||:||:| :||:||||||:||:|:|||:||| || ||||::|
Qy   241 CAYCARYTNAARYTNGAYGTNGTNWSNTAYAAYGAYTTYYTNGGNCARATHATHGAYCAR 300

Db   927 tatacttcagtagctgtaactggtgcacatggtaaaacttctacaacaggtttattatca 986
         |||::|:: || ||||:|||:|||||||||||:||||||:|| |||||||||:||:|||
Qy   301 TAYACNWSNGTNGCNGTNACNGGNGCNCAYGGNAARACNWSNACNACNGGNYTNTNWSN 360

Db   987 catgttatgaatggtgataaaaagacctcatttaattggtgatggcacaggtatggga 1046
         ||:||:|||||:|| |||||| |||:||:|||||:|| || || ||:||:|||||||:|
Qy   361 CAYGTNATGAAYGGNGAYAARAARACNWSNTTYYTNATHGGNGAYGGNACNGGNATGGGN 420

Db  1047 ttgcctgaaagtgattatttcgctttgggcatggaatatagacgtcactttttaagt 1106
         :| |||||:|||||||:|:|| ||:||||:|||||:||:|| ||:||:||:||::|||:
Qy   421 YTNCCNGARWSNGAYTAYTTYGCNTTYGGNCNTTYGARGCNTNGYGARTAYMGNMGNCAYTTYTNWSN 480

Db  1107 tataaacctgattacgcaattatgacaatatattgnnnnnnnnnnnnnnnnnnnnnn 1166
         |||:::|:|| ||:|||:|||:| ||| || |:|:|| ||::: ||:||||||||||
Qy   481 TAYAARCCNGAYTAYGCNATHATGACNAAYATHGAYTTYGAYCAYCCNGAYTAYTTYAAR 540

Db  1167 nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnntggcacataatgttaaaaaggtatt 1226
                                                   --   |  || ||:||:|:|| |||
Qy   541 GAYATHAAYGAYGTNTTYGAYGCNTTYCARGARATGGCNCAYAAYGTNAARAARGGNATH 600

Db  1227 attgctggggtgatgatgaacatctacgtaaaattgaagcagatgttccaatttattat 1286
         ||:|||||| || |||||||||||:|||:| || |||||:|||||:|| |||:|||:|:
Qy   601 ATHGCNTGGGGNGAYGAYGARCAYATYMGNAARAARATHGARGCNGAYGTNCCNATHTAYTAY 660

Db  1287 tatggatttaaagattccggatgctgacatttgtccaaaattaccggataaaggt 1346
         ||:|||::||:||:|| ||:||| |||:||:|| ||:||:|||::|| |||||||:||
Qy   661 TAYGGNTTYAARGAYWSNGAYGAYATHTAYGCNCARAAYATHCARATHACNGAYAARGGN 720

Db  1347 actgctttgatgtatgtgatggatggatggagttttatgatcatcctgtctccacaatat 1406
         ||:||:||| || |||||:||||||:||:|:|| ::| ||||:||:||:||: |||||||:
Qy   721 ACNGCNYTYGAYGTNATGYGAYGGNGARTTYTAYGAYCAYYTNTNWSNCCNCARTAY 780

Db  1407 ggtgaccatacagtttaaatgctaaatgcgtaattgcgattagtttattagagaagcta 1466
         || |||||:|||| |||:||:|||:| ||::|||||||| || ||||:|||||||:|||
Qy   781 GGNGAYCAYACNGTNTNAAYGCNTYNGCNTGNGCNATHAGYTTYATYTNGARAAARYTN 840

Db  1467 gatgttacaaatattaaggaagcattaagaaacgtttggtgttaaacgtcgtttcaat 1526
         ||:|| ||:||| |||:||:||||| ||:|||:|||||:||:|| |||:|||:||:||
```

```
Qy   841 GAYGTNACNAAYATHAARGARGCNYTNGARACNTTYGGNGGNGTNAARMGNMGNTTYAAY 900

Db  1527 gaaactacaattgcaaatcaagttattgtagatgattatgcacaccatccaagagaaat 1586
         ||:|||||:||:||:||:||||| ||:||:||:||:|||||:||:|||||:|| ||:|||
Qy   901 GARACNACNATHGCNAAYCARGTNATHGTNGAYGAYTAYGCNCAYCAYCCNMGNGARATH 960

Db  1587 agtgctacaagacacgaaagaaatatccacataagaagttgtcagtatt 1646
         :|:|| ||:||:|| ||| |||||:|| ||||||||:||:||:||:||::||
Qy   961 WSNGCNACNATHGAYCACNGCNMGNAARAARTAYCCNCAYAARGARGTNGTNGCNGTNTTY 1020

Db  1647 caaccacacactttcctctagaaccacaggccatttaaatgaattgcagaaagtttaagt 1706
         ||:||:||:|| ||:||: | ::| :| ||:||:||:||:||:||:|||| |:||::|:|
Qy  1021 CARCCNCAYACNTTYWSNMGNACNCARGCNTTYYTNAAYGARTTYGCNGARWSNYNTGY 1080

Db  1707 aaagcagatcgtgattcctatgtgaaattttggatcaattagagaaaatactggcgca 1766
         |||||:||:||:||:||:||:||:||:||::||||:||:::|| |||||:|| |||:||
Qy  1081 AARGCNGAYMGNGTNTTYTNTGYGARATHTTYGGNWSNATHMGNGARAAYWSNGGNGCN 1140

Db  1767 ttaacgatacaagatttaattgataaaattgaagtgcatcgttaattaatgaagattct 1826
         :| :|:| ||:|:||::||:|:: :||||| |:|| ||:|:| ||:|||| |:|::
Qy  1141 YTNACNATHCARGAYYTNATHGAYAARATHGGNGNCNWSNTTYATHAAYGARGAYYTN 1200

Db  1827 attaatgattagaacaatttgataatgctgttattttatttatggtgcaggtgatatt 1886
         |||:||:|:|:||:||:||:||:|| ||:| |:|| | ||||:|||| ||||:||||||
Qy  1201 ATHAAYGTNYTNGARCARTTYGAYAAYGCNGTNYTNTTYATGGNGCNGGNGAYATH 1260

Db  1887 caaaattacaaaatgcatatttagataaattaggcatgaaaaatgcgttt 1937
         ||:||:||:|||||:||:||:|:||:|||||:|||:||:|||||||:|||
Qy  1261 CARAARYTNCARAAYGCNTAYYTNGAYAARYTNGGNATGAARAAYGCNTTY 1311
```

```
RESULT   3
ID   V80065 standard; DNA; 660 BP.
AC   V80065;
DT   17-MAR-1999 (first entry)
DE   Partial nucleotide sequence of the MurC gene.
KW   MurC gene; UDP-N-acetylmuramate:L-alanine ligase; MurC polypeptide;
KW   bacterial; infection; H. pylori; cancer; ulcer; gastritis; vaccine;
KW   immunogen; drug; genetic immunisation; ds.
OS   Staphylococcus aureus.
FH   Key         Location/Qualifiers
FT   CDS         2..660
FT               /*tag=  a
FT               /product= "partial MurC polypeptide"
PN   EP-889123-A2.
PD   07-JAN-1999.
PF   26-JUN-1998; 305064.
PR   03-JUL-1997; US-052720.
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PA   (SMIK ) SMITHKLINE BEECHAM PLC.
PI   Burnham MKR, Wallis NG;
DR   WPI; 99-062655/06.
DR   P-PSDB; W89199.
PT   New isolated MurC polypeptide from Staphylococcus aureus and related
PT   nucleic acid - useful in diagnosis, treatment and prevention of
PT   bacterial infections
PS   Claim 2; Pages 4-5; 39pp; English.
CC   The invention relates to a UDP-N-acetylmuramate:L-alanine ligase
CC   (MurC polypeptide) encoded by the S. aureus MurC gene. Host cells
CC   containing an expression system comprising the MurC gene can be used for
CC   the recombinant production of the polypeptide. Agonists or the MurC
CC   polypeptide are used to treat conditions requiring increased activity or
CC   expression of the polypeptide. Antagonists, inhibitory nucleic acid or
CC   competitive polypeptide are useful for inhibiting the polypeptide e.g.
CC   bacterial (especially S. aureus) infections. They are also useful against
CC   Helicobacter pylori infections and related cancers, ulcers and gastritis.
CC   The antibacterial agents are useful to treat in-dwelling devices for
CC   infection prevention or generally as wound treatments to prevent adhesion
CC   of bacteria to matrix proteins. The MurC polypeptide is also useful for
CC   diagnosing or prognosing a (susceptibility to) disease, for raising
CC   antibodies; to identify modulators or specific receptors; in rational
CC   drug design and as an immunogen for vaccines. The MurC gene sequences are
CC   useful in antisense/ribozyme therapeutics; to detect mutant MurC gene;
```

```
 __   __  _____   _____  ____    _____  _   _
|  \ /  ||  _  | / ____||  _ \  / ____|| | | |
|   V   || |_| || (___  | |_) || |     | |_| |
| |\_/| ||  ___| \___ \ |  _ < | |     |  _  |
| |   | || |     ____) || |_) || |____ | | | |  (TM)
|_|   |_||_|    |_____/ |____/  \_____||_| |_|
```

```
**************************************************

**************************************************
```

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nn   n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Sat Nov 27 12:32:46 1999;   MasPar time 162.98 Seconds
                      867.701 Million cell updates/sec

Tabular output not generated.

Title: >US-09-103-287-3
Description: (1-660) from US09103287.seq
Perfect Score: 660
N.A. Sequence: 1 ATTAAAGATTCGGATGACA..........GCGTTTAATATGTTTATAA 660
Comp: TAAATTTCTAAGCCTACTGT..........CGCAAAATTATACAAATATT

Scoring table: TABLE default
              Gap 6

Nmatch   STD :   Dbase 0;   Query 0

Searched: 271905 seqs, 107135622 bases x 2

Post-processing: Minimum Match 0%
                  Listing first 45 summaries

Database: n-geneseq35
             1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
             8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
             14:part14 15:part15 16:part16 17:part17 18:part18
             19:part19 20:part20 21:part21 22:part22 23:part23
             24:part24 25:part25 26:part26 27:part27 28:part28
             29:part29 30:part30 31:part31 32:part32 33:part33
             34:part34 35:part35 36:part36 37:part37 38:part38
             39:part39 40:part40 41:part41 42:part42 43:part43
             44:part44 45:part45 46:part46 47:part47 48:part48
             49:part49 50:part50 51:part51 52:part52 53:part53
             54:part54 55:part55 56:part56 57:part57 58:part58
             59:part59 60:part60

Statistics: Mean 8.484;   Variance 6.123;   scale 1.386

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Score	% Query Match	Length	DB ID	ID	Description	Pred. No.
1	646	97.9	2424	58	V74703	Staphylococcus aureus	0.00e+00
2	638	96.7	660	60	V80065	Partial nucleotide se	0.00e+00
3	638	96.7	1351	60	V99650	UDP-N-acetylmuramate:	0.00e+00
4	587	88.9	619	48	V53479	DNA encoding a Staphy	0.00e+00
5	111	16.8	677	60	X13717	Enterococcus faecalis	1.41e-45
6	92	13.9	1267	46	V27381	Streptococcus pneumon	5.16e-35
7	92	13.9	11864	45	V52194	Streptococcus pneumon	5.16e-35
c 8	80	12.1	1825	47	V43027	Streptococcus pneumon	1.94e-28
c 9	38	5.8	91	9	Q51746	Oligonucleotide probe	1.37e-06

c	10	37	5.6	204	1	N81164	Base substituted E.co	4.14e-06
	11	36	5.5	204	1	N81164	Base substituted E.co	1.24e-05
	12	34	5.2	91	9	Q51746	Oligonucleotide probe	1.09e-04
c	13	34	5.2	114	12	Q70466	Generic DNA sequence	1.09e-04
	14	33	5.0	91	46	V44650	Mammalian DNA replica	3.18e-04
c	15	32	4.8	91	46	V44650	Mammalian DNA replica	9.22e-04
c	16	32	4.8	114	12	Q70468	Generic DNA sequence	9.22e-04
c	17	32	4.8	114	12	Q70465	Generic DNA sequence	9.22e-04
c	18	32	4.8	114	12	Q70469	Generic DNA sequence	9.22e-04
c	19	26	3.9	114	12	Q70467	Generic DNA sequence	4.33e-01
c	20	26	3.9	114	12	Q70468	Generic DNA sequence	4.33e-01
	21	26	3.9	498	3	N50034	Sequence encoding new	4.33e-01
	22	26	3.9	501	3	N50032	Sequence encoding new	4.33e-01
	23	26	3.9	501	3	N50029	Sequence encoding new	4.33e-01
	24	26	3.9	501	3	N50028	Sequence encoding new	4.33e-01
	25	26	3.9	501	3	N50026	Sequence encoding new	4.33e-01
	26	26	3.9	501	3	N50030	Sequence encoding new	4.33e-01
	27	26	3.9	501	3	N50024	Sequence encoding new	4.33e-01
	28	26	3.9	501	3	N50033	Sequence encoding new	4.33e-01
	29	26	3.9	501	3	N50031	Sequence encoding new	4.33e-01
	30	25	3.8	114	12	Q70465	Generic DNA sequence	1.15e+00
	31	25	3.8	114	12	Q70467	Generic DNA sequence	1.15e+00
	32	25	3.8	114	12	Q70469	Generic DNA sequence	1.15e+00
c	33	25	3.8	501	3	N50027	Sequence encoding new	1.15e+00
c	34	25	3.8	501	3	N50030	Sequence encoding new	1.15e+00
c	35	25	3.8	501	3	N50025	Sequence encoding new	1.15e+00
c	36	25	3.8	1200	2	Q10187	Sequence encoding fun	1.15e+00
c	37	24	3.6	39	7	Q51787	Mixed oligonucleotide	3.02e+00
c	38	24	3.6	108	23	T29064	Probe for Candida alb	3.02e+00
c	39	24	3.6	114	12	Q70470	Generic DNA sequence	3.02e+00
c	40	24	3.6	146	21	T13624	ME#2b library generat	3.02e+00
c	41	24	3.6	1440	26	T28571	Bacterial antibiotic	3.02e+00
c	42	24	3.6	1676	60	X13557	Enterococcus faecalis	3.02e+00
	43	24	3.6	2171	1	N81668	Sequence of part of t	3.02e+00
	44	24	3.6	2403	21	T12562	S.aureus topoisomeras	3.02e+00
	45	24	3.6	2403	21	T12573	S.aureus mutant grlA	3.02e+00

ALIGNMENTS

RESULT   1
ID   V74703 standard; DNA; 2424 BP.
AC   V74703;
DT   16-MAR-1999   (first entry)
DE   Staphylococcus aureus contig SEQ ID #392
KW   Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW   cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW   skin infection; surgical wound infection; scalded skin syndrome;
KW   toxic shock syndrome; ds.
OS   Staphylococcus aureus.
FH   Key   Location/Qualifiers
FT   misc_feature   1141..1200
FT   /*tag= a
FT   /note= "these bases represent a line of missing text in
FT   the sequence listing in the specification. They
FT   are included to maintain the nucleotide numbering
FT   given in the specification for this DNA sequence
PN   EP-786519-A2.
PD   30-JUL-1997.
PF   07-JAN-1997; 100117.
PR   05-JAN-1996; US-009861.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI   Rosen CA;
DR   WPI; 97-374922/35.
PT   Polynucleotide(s) and proteins derived from Staphylococcus aureus -
PT   stored on computer readable medium and used in the production of
PT   anti-S.aureus vaccines
PS   Claim 1; Page 1287-1288; 3271pp; English.
CC   This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC   of the invention. The DNA sequences are recorded on a computer readable
CC   medium, preferably selected from a floppy or hard disk, random access
CC   memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
```

CC   the S.aureus DNA sequences allows putative functions to be assigned so
CC   that protein-encoding or regulatory regions of commercial, therapeutic or
CC   industrial importance can be obtained. Specifically, sequences which are
CC   likely to encode antigens have been identified and these polypeptides can
CC   be used in a vaccine composition against S.aureus infection. The
CC   polypeptides can also be used in a kit for the immunodetection of
CC   S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC   including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC   skin and surgical wound infection, scalded skin syndrome, toxic shock
CC   syndrome, etc. Organisms transformed with the DNA sequences can be used
CC   for recombinant production of the polypeptides. The new DNA sequences
CC   (and their fragments) are useful as primers or probes for isolating
CC   homologues of any of the S.aureus DNA sequences contained on the
CC   computer readable medium.
SQ   Sequence   2424 BP;   839 A;   302 C;   462 G;   759 T;

Query Match
Best Local Similarity   97.9%;   Score 646;   DB 58;   Length 2424;
Matches   653;   Conservative   0;   Mismatches   7;   Indels   0;   Gaps   0;

Db   1292  atttaaagattcggatgacatttatgtcaaaatattcaaattacggataaagtactgc 1351
            ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Qy      1  ATTTAAAGATTCGGATGACATTTATGTCAAAATATTCAAATATTCAAATTACGGATAAAGTACTGC  60

Db   1352  tttgatgtgtatgtggatggtgagtttatgatcacttcctgtctccacaatatggtga 1411
            | |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     61  TGTTGATGTGTATGTGGATGGTGAGTTTTATGATCACTTCCTGTCTCCACAATATGGTGA 120

Db   1412  ccatacagtttaaatgcattagctgtgattagttcgattagtagttattagagaagctagagatgt 1471
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    121  CCATACAGTTTAAATGCATTAGCTGTGATTAGTTCGATTAGTAGTTATTAGAGAAGCTAGAGATGT 180

Db   1472  tacaaatattaagaagcattagaaacgttgtggtgttaaacgtcgtttcaatgaaac 1531
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    181  TACAAATATTAAAGAAGCATTAGAAACGTTTGGTGGTGTTAAACGTCGTTTCAATGAAAC 240

Db   1532  tacaattgcaaatcaagttatgtagatgatatgcacaccatccaagagaaattagtgc 1591
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    241  TACAATTGCAAATCAAGTTATGTGAGATGATTATGCACACCATCCAAGAGAAATTAGTGC 300

Db   1592  tacaattgaaacagcacgaagaaatatccacataaagaagttgttcagtatttcaacc 1651
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    301  TACAATTGACACAGCACCAGCACGAAAGAAATATCCACATAAAGAAGTTGTTGCAGTATTTCAACC 360

Db   1652  acacacttctctagaacacaggcatttttaaatgatttgcagaaagtttaagtaaagc 1711
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    361  ACACACTTTCTCTAGAACACAAGCACATTTTAAATGATTTGCAGAAAGTTTAAGTAAAGC 420

Db   1712  agatcgtatcctatgtgaaatttttggatcaattagagaaaaatactggcgcattaac 1771
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    421  AGATCGTGTATTCTTATGTGAAATTTTTGGATCAATTAGAGAAAAATACTGGCGCATTAAC 480

Db   1772  gatacaagatttaattgataaaattgaagtgcatcgtaattaatgaagattctattaa 1831
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    481  GATACAAGATTTAATTGATAAAATTGAAGGTGCATCGTAATTAATGAAGATTCTATTAA 540

Db   1832  tgtattagacaatttgatatgctgtattttattatggtgcaggtgatattcaaaa 1891
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    541  TGTATTAGAACAATTTGATAATTGCTGTGTTTTATTTATGGGGTGCAGGTGATATTCAAAA 600

Db   1892  agatcgtatcttagatacatattaggcatgaaaaatgcgtttaaatatgtttataa 1951
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    601  ATTACAAAATGCATATTTAGATAAAATTAGATAAAATAGGCATGAAAAATGCGTTTAAATATGTTTATAA 660

RESULT   2
ID   V80065   standard; DNA; 660 BP.
AC   V80065;
DT   17-MAR-1999   (first entry)
DE   Partial nucleotide sequence of the MurC gene.
KW   MurC gene; UDP-N-acetylmuramate:L-alanine ligase; MurC polypeptide;
KW   bacterial; infection; H. pylori; cancer; ulcer; gastritis; vaccine;
KW   immunogen; drug; genetic immunisation; ds.

OS   Staphylococcus aureus.
FH   Key             Location/Qualifiers
FT   CDS             2..660
FT                   /tag= a
FT                   /product= "partial MurC polypeptide"
PN   EP-889123-A2.
PD   07-JAN-1999.
PF   26-JUN-1998; 305064.
PR   03-JUL-1997; US-052720.
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PA   (SMIK ) SMITHKLINE BEECHAM PLC.
PI   Burnham MKR, Wallis NG;
DR   WPI; 99-062655/06.
DR   P-PSDB; W89199.
PT   New isolated MurC polypeptide from Staphylococcus aureus and related
PT   nucleic acid - useful in diagnosis, treatment and prevention of
PT   bacterial infections
PS   Claim 2; Pages 4-5; 39pp; English.
CC   The invention relates to a UDP-N-acetylmuramate:L-alanine ligase
CC   (MurC polypeptide) encoded by the S. aureus MurC gene. Host cells
CC   containing an expression system comprising the MurC gene can be used for
CC   the recombinant production of the polypeptide. Agonists or the MurC
CC   polypeptide are used to treat conditions requiring increased activity or
CC   expression of the polypeptide. Antagonists, inhibitory nucleic acid or
CC   competitive polypeptide are useful for inhibiting the polypeptide e.g.
CC   bacterial (especially S. aureus) infections. They are also useful against
CC   Helicobacter pylori infections and related cancers, ulcers and gastritis.
CC   The antibacterial agents are useful to treat in-dwelling devices for
CC   infection prevention or generally as wound treatments to prevent adhesion
CC   of bacteria to matrix proteins. The MurC polypeptide is also useful for
CC   diagnosing or prognosing a (susceptibility to) disease, for raising
CC   antibodies; to identify modulators or specific receptors; in rational
CC   drug design and as an immunogen for vaccines. The MurC gene sequences are
CC   useful in antisense/ribozyme therapeutics; to detect mutant MurC gene;
CC   for chromosomal mapping; to determine bacterial serotype; and for genetic
CC   immunisation. The present sequence represents a partial nucleotide
CC   sequence (MurC ORF) of the MurC gene.
SQ   Sequence   660 BP;   233 A;   83 C;   120 G;   224 T;

Query Match
Best Local Similarity   96.7%;   Score 638;   DB 60;   Length 660;
Matches   649;   Conservative   0;   Mismatches   11;   Indels   0;   Gaps   0;

Db      1  atttaaagattcggatgacatttatgtcaaaatattcaaattacggataaagtactgc  60
            ||||||||||||||||||||||||||||| ||||||||| ||||||||||||||||||||
Qy      1  ATTTAAAGATTCGGATGACATTTATGTCAAAATATTCAAATATTCAAATTACGGATAAAGTACTGC  60

Db     61  tttgatgtgtatgtggatggtgagtttatgatcacttcctgtctccacaatatggtga 120
            | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     61  TGTTGATGTGTATGTGGATGGTGAGTTTTATGATCACTTCCTGTCTCCACAATATGGTGA 120

Db    121  ccatacagtttaaatgcattagctgtgattagttcgattagtagttattagagaagctagagatgt 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    121  CCATACAGTTTAAAATGCATTAGCTGTGATTAGTTCGATTAGTAGTTATTAGAGAAGCTAGAGATGT 180

Db    181  tacaaatattaagaagcattagaaacgtttggtggtgttaaacgtcgtttcaatgaaac 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    181  TACAAATATTAAAGAAGCATTAGAAACGTTTGGTGGTGTTAAACGTCGTTTCAATGAAAC 240

Db    241  tacaattgcaaatcaagttatgtagatgattatgcacaccatccaagagaaattagtgc 300
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    241  TACAATTGCAAATCAAGTTATGTTATGATCAAGTTATGTAGATGATTATGAGAAAATAGTGC 300

Db    301  tacaattgacacagcacgaagaaatatccacataaagaagttgttgcagtatttcaacc 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    301  TACAATTGACACAGCACCAGCACGAAAGAAAATATCCACATAAAGAAGTTGTTGCAGTATTTCAACC 360

Db    361  acacacttctctagaacacaggcatttttaaatgatttgcagaaagtttatgtaaagc 420
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    361  ACACACTTTCTCTAGAACACAAGCACATTTTAAATGATTTTAAATGATTTTGCAGAAAGTTTAAGTAAAGC 420

Db    421  agatcgtgtattcttatgtgaaatttttggctcaattagagaaaatcctggcgcattaac 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

```
 __   __  ___   ___   ___   ___   _   _
|  \ /  ||  _ \ / __| |  _ \ / __| | | | |
| |\V/| || |_) |\__ \ | |_) |\__ \ | |_| |
|_|   |_||  __/ |___/ |_| \_\|___/ |_| |_|   (TM)
          |_|
```

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_tpn n.a. - n.a. Smith-Waterman search, using a protein query
            which has been backtranslated into n.a. using IUPAC symbols

Perfect Score: 2663

N.A. Sequence: 1 TTYAARGAYWSNGAYGAYAT.........TNGGNATGAARAAYGCNYTY 645
                   AARTTYCTRWSNCTRCTRTA.........ANCCNTACTYYTRCGNAAR

Comp:

Run on: Sat Nov 27 10:20:03 1999;   MasPar time 159.66 Seconds
                                            865.603 Million cell updates/sec

Tabular output not generated.

Title: >US-09-103-287-4
Description: (1-215) from US09103287.pep

Scoring table: TABLE bktranslate2
                Gap 30

Nmatch STD :   Dbase 0; Query 0

Searched: 271905 seqs, 107135622 bases x 2

Post-processing: Minimum Match 0%
                 Listing first 45 summaries

Database: n-geneseq35
                1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
                8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
                14:part14 15:part15 16:part16 17:part17 18:part18
                19:part19 20:part20 21:part21 22:part22 23:part23
                24:part24 25:part25 26:part26 27:part27 28:part28
                29:part29 30:part30 31:part31 32:part32 33:part33
                34:part34 35:part35 36:part36 37:part37 38:part38
                39:part39 40:part40 41:part41 42:part42 43:part43
                44:part44 45:part45 46:part46 47:part47 48:part48
                49:part49 50:part50 51:part51 52:part52 53:part53
                54:part54 55:part55 56:part56 57:part57 58:part58
                59:part59 60:part60
```

Statistics: Mean 50.914; Variance 239.321; scale 0.213

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed.
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	2573	96.6	2424	58	V74703	Staphylococcus aureus	3.53e-212
2	2557	96.0	660	60	V80065	Partial nucleotide se	9.99e-211
3	2557	96.0	1351	60	V99650	UDP-N-acetylmuramate:	9.99e-211
4 c	2329	87.5	619	48	V53479	DNA encoding a Staphy	4.65e-190
5	934	35.1	677	60	X13717	Enterococcus faecalis	4.66e-65
6 c	749	28.1	1267	46	V27381	Streptococcus pneumon	6.67e-49
7 c	749	28.1	11864	47	V52194	Streptococcus pneumon	6.67e-49
8 c	699	26.2	1825	47	V43027	Streptococcus pneumon	1.42e-44

9	222	8.3	861	30	T77403	H. pylori cytoplasmic	2.48e-05
10	222	8.3	861	30	T67723	H. pylori cytoplasmic	2.48e-05
11	202	7.6	1365	30	T67859	H. pylori cytoplasmic	7.08e-04
12	183	6.9	843	47	V52489	Streptococcus pneumon	1.57e-02
13	183	6.9	1003	46	V27378	Streptococcus pneumon	1.57e-02
14	183	6.9	1350	36	T89154	Streptococcus pneumon	1.57e-02
15	183	6.9	1350	52	V33971	S. pneumoniae MurD mR	1.57e-02
16	183	6.9	1353	52	V33970	S. pneumoniae MurD ge	1.57e-02
17	183	6.9	6217	50	V65252	DNA encoding S. pneum	1.57e-02
18 c	183	6.9	6693	47	V52328	Streptococcus pneumon	1.57e-02
19	152	5.7	1532	47	V59408	Synthetic HIV gag gen	2.30e+00
20	151	5.7	534	52	V45173	Degenerate nucleotide	1.98e+00
21	140	5.3	501	3	N50026	Sequence encoding new	1.17e+01
22	140	5.3	501	3	N50024	Sequence encoding new	1.17e+01
23 c	141	5.3	1311	59	V81395	Degenerate human tumo	1.01e+01
24 c	138	5.2	2193	60	V84038	Human gelsolin gene c	1.56e+01
25	139	5.2	8247	3	N70996	Sequence encoding N-t	1.35e+01
26 c	135	5.1	1026	30	T68253	H. pylori cytoplasmic	2.41e+01
27 c	135	5.1	1071	48	V19280	Pseudomonas fluoresce	2.41e+01
28	137	5.1	1782	1	Q03856	Heat shock protein (H	1.81e+01
29 c	133	5.0	501	3	N50024	Sequence encoding new	3.21e+01
30 c	134	5.0	510	42	V28775	Human interleukin-XX	2.78e+01
31 c	134	5.0	984	17	Q94336	Degenerate Alteromona	2.78e+01
32	132	5.0	1640	2	N60590	Sequence encoding mat	3.70e+01
33	132	5.0	1641	17	T08561	Human CSF-1 cDNA clon	3.70e+01
34	132	5.0	1641	60	X13057	Enterococcus faecalis	3.70e+01
35	132	5.0	1642	2	N80775	Sequence of a cDNA en	3.70e+01
36	132	5.0	1642	1	N91084	Colony stimulating fa	3.70e+01
37	132	5.0	1642	37	V01080	Gene for human colony	3.70e+01
38	132	5.0	1642	33	T80591	Human colony stimulat	3.70e+01
39	132	5.0	1642	26	T58753	Human short form CSF-	3.70e+01
40	132	5.0	1880	46	V35031	Arabidopsis thaliana	3.70e+01
41	132	5.0	2302	35	T94747	Human colony stimulat	3.70e+01
42 c	132	5.0	2309	13	Q73786	Partial PCR fragment	3.70e+01
43	132	5.0	2500	7	Q43710	Sequence of clone lam	3.70e+01
44 c	132	5.0	4315	60	X12938	Enterococcus faecalis	3.70e+01
45 c	134	5.0	9767	60	X13067	Enterococcus faecalis	2.78e+01

ALIGNMENTS

```
RESULT 1
ID   V74703 standard; DNA; 2424 BP.
AC   V74703;
DT   16-MAR-1999  (first entry)
DE   Staphylococcus aureus contig SEQ ID #392.
KW   Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW   cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW   skin infection; surgical wound infection; scalded skin syndrome;
OS   toxic shock syndrome; ds.
OS   Staphylococcus aureus.
FH   Key             Location/Qualifiers
FT   misc_feature    1141..1200
FT                   /*tag= a
FT                   /note= "these bases represent a line of missing text in
FT                   the sequence listing in the specification. They
FT                   are included to maintain the nucleotide numbering
FT                   given in the specification for this DNA sequence"
PN   EP-786519-A2.
PD   30-JUL-1997;
PF   07-JAN-1997; 100117.
PR   05-JAN-1996; US-009861.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI   Rosen CA;
DR   WPI; 97-374922/35.
PT   Polynucleotide(s) and proteins derived from Staphylococcus aureus
PT   stored on computer readable medium and used in the production of
PT   anti-S.aureus vaccines
PS   Claim 1; Page 1287-1288; 3271pp; English.
CC   This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC   of the invention. The DNA sequences are recorded on a computer readable
CC   medium, preferably selected from a floppy or hard disk, random access
```

CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC the S.aureus DNA sequences allows putative functions to be assigned so
CC that protein-encoding or regulatory regions of commercial, therapeutic or
CC industrial importance can be obtained. Specifically, sequences which are
CC likely to encode antigens have been identified and these polypeptides can
CC be used in a vaccine composition against S.aureus infection. The
CC polypeptides can also be used in a kit for the immunodetection of
CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC skin and surgical wound infections, scalded skin syndrome, toxic shock
CC syndrome, etc. Organisms transformed with the DNA sequences can be used
CC for recombinant production of the polypeptides. The new DNA sequences
CC (and their fragments) are useful as primers or probes for isolating
CC homologues of any of the S.aureus DNA sequences contained on the
CC computer readable medium.
SQ Sequence 2424 BP; 839 A; 302 C; 462 G; 759 T;

Query Match 96.6%; Score 2573; DB 58; Length 2424;
Best Local Similarity 59.5%; Pred. No. 3.53e-212;
Matches 384; Conservative 158; Mismatches 103; Indels 0; Gaps 0;

```
Db 1293 tttaaagattcggatgacattatgctcaaaatattccaaattacggataaagtactgct 1352
           ||:|:||  ||:|:||:|| ::|| :::|:::|:||:|::|  | |:|||:||:|| ||
Qy    1 TTYAARGAYWSNGAYGAYATHTAYGCNCARATHTTYCARATHACNGAYAARGGNACNGCN 60

Db 1353 tttgatgtgtatggtggtgagtttatgatcacttcctgtctccacaaatatggtgac 1412
           ||:|| ||:|| ||:|| || ||:||||:||:||:||  ||:|||| || ||:||
Qy   61 GTNGAYGTNTAYGTNGAYGGNGARTTYATGAYCAYTTYYTNWSNCCNCARTAYGGNGAY 120

Db 1413 catacagtttaaatgcattagctgtaattgcgtattgttattagagaagctagatgtt 1472
           ||:||:|| :|:|| ||:|| || ||:||:|| ||:||:|| |||||:||:|| ||
Qy  121 CAYACNGTNYTNAAYGCNYTNGCNGTNATHGCNATHWSNTAYYTNGARAARYTNGAYGTN 180

Db 1473 acaaatattaaagaagcattagaaacgtttggtggtgttaaacgtcgtttcaatgaaact 1532
           || ||:||:||:||:|| ||:|| || ||:|| ||:||:||:||:||:|| ||:||:||
Qy  181 ACNAAYATHAARGARGCNYTNGARACNTTYGGNGGNGTNAARMGNMGNTTYAAYGARACN 240

Db 1533 acaattgcaaatcaagttattgtagatgattatgcaccaccatccaagagaaattagtgct 1592
           || ||:|| || || ||:||:||:||:|| || ||:|| || ||:||:||:||:|| ||
Qy  241 ACNATHGCNAAYCARGTNATHGTNGAYGAYTAYGCNCAYCCNMGNGARATHWSNGCN 300

Db 1593 acaattgaaacagcacgcgaaagaaatatccacacacaagaagttgttgcagtatttcaacca 1652
           || ||:|| || ||:|| || || || ||:|| ||:||:||:|| ||:|| ||:||:|| ||
Qy  301 ACNATHGAYACNGCNMGNGAAARARTAYCCNCAYACNAARGARGTNGNCGNTNTTYCARCCN 360

Db 1653 cacactttctctcagaacacagcatttttaatgaatttgcgaaagtttaagtaaagca 1712
           || ||:||:||:||:|| || ||:||:|| ||:||:|| ||:|| ||:||:||:|| ||
Qy  361 CAYACNTTYWSNMGNACNCARGCNTTYYTNAAYGARTTYGCNGARWSNYTNWSNAARGCN 420

Db 1713 gatcggtgattccttatgtgaaatttttggatcaattagaagaaatactggcgcattaacg 1772
           ||:|| ||:||:|| ||:||:|| ||:|| ||:||:|| || || ||:|| ||:||:||
Qy  421 GAYMGNGTNTTYYTNTGYGARATHTTYGGNWSNATHMGNGARAAYACNGGNGCNYTNACN 480

Db 1773 atacaagatttaattgataaaattgaagtgcatcgtgttaattaatgaagattctattaat 1832
           || || ||:||:|| ||:||:||:|| || ||:||:||:|| ||:|| ||:||:|| ||
Qy  481 ATHCARGAYYTNATHGAYAARATHGARGGNGCNWSNYTNATHAAYGARGAYWSNATHAAY 540

Db 1833 gtattagaacatttgataatgctgttatttatttatggtgcaggtgatattcaaaaa 1892
           ||:||:|| ||:||:|| ||:|| ||:||:||:|| ||:|| ||:|| ||:|| ||
Qy  541 GTNYTNGARCARTTYGAYAAYGCNGTNGTNYTNTATGGGNCNGGNGAYATHCARAAR 600

Db 1893 ttacaaaatgcatattagataaattaggcatgaaaatgcgttt 1937
           ||:|| ||:|| ||:|| ||:||:|| ||:|| || ||:||
Qy  601 YTNCARAAYGCNTAYYTNGAYAAYGCNTTYYTNGGNATGAARAAYGCNTTY 645
```

RESULT 2
ID V80065 standard; DNA; 660 BP.
AC V80065;
DT 17-MAR-1999 (first entry)
DE Partial nucleotide sequence of the MurC gene.
KW MurC gene; UDP-N-acetylmuramate:L-alanine ligase; MurC polypeptide;

CC The invention relates to a UDP-N-acetylmuramate:L-alanine ligase
CC (MurC polypeptide) encoded by the S. aureus MurC gene. Host cells
CC containing an expression system comprising the MurC gene can be used for
CC the recombinant production of the polypeptide. Agonists or the MurC
CC polypeptide are used to treat conditions requiring increased activity or
CC expression of the polypeptide. Antagonists, inhibitory nucleic acid or
CC competitive polypeptide are useful for inhibiting the polypeptide e.g.
CC bacterial (especially S. aureus) infections. They are also useful against
CC Helicobacter pylori infections and related cancers, ulcers and gastritis.
CC The antibacterial agents are useful to treat in-dwelling devices for
CC infection prevention or generally as wound treatments to prevent adhesion
CC of bacteria to matrix proteins. The MurC polypeptide is also useful for
CC diagnosing or prognosing a (susceptibility to) disease, for raising
CC antibodies; to identify modulators or specific receptors; in rational
CC drug design and as an immunogen for vaccines. The MurC gene sequences are
CC useful in antisense/ribozyme therapeutics; to detect mutant MurC gene;
CC for chromosomal mapping; to determine bacterial serotype; and for genetic
CC immunisation. The present sequence represents a partial nucleotide
CC sequence (MurC ORF) of the MurC gene.
SQ Sequence 660 BP; 233 A; 83 C; 120 G; 224 T;

Query Match 96.0%; Score 2557; DB 60; Length 660;
Best Local Similarity 59.4%; Pred. No. 9.99e-211;
Matches 383; Conservative 157; Mismatches 105; Indels 0; Gaps 0;

```
Db    2 tttaaagattcggatgacattatgctcaaaatattcaaattacggataaagtactgct 61
           ||:|:||  ||:|:||:|| ::|| :::|:::|:||:|::| | |:|||:||:|| ||
Qy    1 TTYAARGAYWSNGAYGAYATHTAYGCNCARATHTTYCARATHACNGAYAARGGNACNGCN 60

Db   62 tttgatgtgtatggtggtgagtttatgatcacttcctgtctccacaaatatggtgac 121
           ||:|| ||:|| ||:|| || ||:||||:||:||:||  ||:|||| || ||:||
Qy   61 GTNGAYGTNTAYGTNGAYGGNGARTTYATGAYCAYTTYYTNWSNCCNCARTAYGGNGAY 120

Db  122 catacagtttaaatgcattagctgtaattgcgtattgttatttattagagaagctagatgtt 181
           ||:||:|| :|:|| ||:|| || ||:||:|| ||:||:|| |||||:||:|| ||
Qy  121 CAYACNGTNYTNAAYGCNYTNGCNGTNATHGCNATHWSNTAYYTNGARAARYTNGAYGTN 180

Db  182 acaaatattaaagaagcattagaaacgtttggtggtgttaaacgtcgtttcaatgaaact 241
           || ||:||:||:||:|| ||:|| || ||:|| ||:||:||:||:||:|| ||:||:||
Qy  181 ACNAAYATHAARGARGCNYTNGARACNTTYGGNGGNGTNAARMGNMGNTTYAAYGARACN 240

Db  242 acaattgcaaatcaagttattgtagatgattatgcacaccatccaagagaaattagtgct 301
           || ||:|| || || ||:||:||:||:|| || ||:|| || ||:||:||:||:|| ||
Qy  241 ACNATHGCNAAYCARGTNATHGTNGAYGAYTAYGCNCAYCCNMGNGARATHWSNGCN 300

Db  302 acaattgacaacagcacgaaagaaatatccacataaagaagtgttgcagtatttccaacca 361
           || ||:|| || ||:|| || || || ||:|| ||:||:||:|| ||:|| ||:|| ||
Qy  301 ACNATHGAYACNGCNMGNAAAARARTAYCCNCAYACNAARGARGTNGCNGTNTTYCARCCN 360

Db  362 cacactttctctagaacaacaagcatttttaaatgaatttgcagaaagtttatgtaaagca 421
           || ||:||:||:||:|| || ||:||:|| ||:||:|| ||:|| ||:||:||:|| ||
Qy  361 CAYACNTTYWSNMGNACNCARGCNTTYYTNAAYGARTTYGCNGARWSNYTNWSNAARGCN 420
```

```
Qy  742  ACTGCTTTGATGTGTATGTGGATGGTGAGTTTTATGATCACTTCCTGTCTCCACAAATAT  801
              T  A  F  D  V  V  Y  V  V  D  G  E  F  Y  D  H  F  L  S  P  Q  Y
Qt         T  A  F  D  V  V  Y  V  V  D  G  E  F  Y  D  H  F  L  S  P  Q  Y

Db  781  GGNGAYCAYACNGTNYTNAAYGCNTYNGCNGTNATHGCNATHWSNATAYTTNGARAARYTN  840
           G  D  H  T  V  L  N  A  L  A  V  I  A  I  S  Y  L  E  K  L
Dt         G  D  H  T  V  L  N  A  L  A  V  I  A  I  S  Y  L  E  K  L

Qy  802  GGTGACCATACAGTTTTAAATGCCGATTAGTTATTGCGATTAGTTATTAGAGAAGCTA  861
           G  D  H  T  V  L  N  A  L  A  V  I  A  I  S  Y  L  E  K  L
Qt         G  D  H  T  V  L  N  A  L  A  V  I  A  I  S  Y  L  E  K  L

Db       GAYGTNACNAAYATHAARGARGCNYTNGARACNTTYGGNGNGTNAARMGNMGNTTYAAY
           D  V  T  N  I  K  E  A  L  E  T  F  G  G  V  K  R  R  F  N
Dt         D  V  T  N  I  K  E  A  L  E  T  F  G  G  V  K  R  R  F  N

Qy  841  GATGTTACAAATATTAAAGAAGCATTAGAAACGTTTGGTGGTGTTAAACGTCGTTTCAAT  900
           D  V  T  N  I  K  E  A  L  E  T  F  G  G  V  K  R  R  F  N
Qt         D  V  T  N  I  K  E  A  L  E  T  F  G  G  V  K  R  R  F  N

Db  901  GARACNACNATHGCNAAYCARGTNATHGTNGAYGAYTAYGCNCAYCCNMGNGARATH  960
           E  T  T  I  A  N  Q  V  I  V  D  D  Y  A  H  H  P  R  E  I
Dt         E  T  T  I  A  N  Q  V  I  V  D  D  Y  A  H  H  P  R  E  I

Qy  922  GAAACTACAATTGCAAATCAAGTTATTGTAGATGATTATGCACACCATCCAAGAGAAATT  981
           E  T  T  I  A  N  Q  V  I  V  D  D  Y  A  H  H  P  R  E  I
Qt         E  T  T  I  A  N  Q  V  I  V  D  D  Y  A  H  H  P  R  E  I

Db  961  WSNGCNACNATHGARACNGCNMGNAAARARTAYCCNCAYGARGTNGTNGCNGTNTTY  1020
           S  A  T  I  E  T  A  R  K  K  Y  P  H  K  E  V  V  A  V  F
Dt         S  A  T  I  E  T  A  R  K  K  Y  P  H  K  E  V  V  A  V  F

Qy  982  AGTGCTACAATTGACACAGCACGAGAAAAATATCCACATAAAGAAGTTGTTGCAGTATTT  1041
           S  A  T  I  D  T  A  R  K  K  Y  P  H  K  E  V  V  A  V  F
Qt         S  A  T  I  D  T  A  R  K  K  Y  P  H  K  E  V  V  A  V  F

Db 1021  CARCCNCAYACNTTYWSNMGNACNCARGCNTTYTNAAYGARTTYGCNGARWSNYTNWSN  1080
           Q  P  H  T  F  S  R  T  Q  A  F  L  N  E  F  A  E  S  L  S
Dt         Q  P  H  T  F  S  R  T  Q  A  F  L  N  E  F  A  E  S  L  S

Qy 1042  CAACCACACTTTCTCTAGAACACAAGCATTTTAAATGAATTGCAGAAAGTTTATGT  1101
           Q  P  H  T  F  S  R  T  Q  A  F  L  N  E  F  A  E  S  L  C
Qt         Q  P  H  T  F  S  R  T  Q  A  F  L  N  E  F  A  E  S  L  C

Db 1081  AARGCNGAYMGNGTNTTYYTNTGYGARATHTTYGGNWSNATHMGNGARAAYACNGGNGCN  1140
           K  A  D  R  V  F  L  C  E  I  F  G  S  I  R  E  N  T  G  A
Dt         K  A  D  R  V  F  L  C  E  I  F  G  S  I  R  E  N  T  G  A

Qy 1102  AAAGCAGATCGTGTATTCTTATGTGAAAATTTTGGCTCATTAGAGAAAATTCTGGCGCA  1161
           K  A  D  R  V  F  L  C  E  I  F  G  S  I  R  E  N  S  G  A
Qt         K  A  D  R  V  F  L  C  E  I  F  G  S  I  R  E  N  S  G  A

Db 1141  YTNACNAYHCARGAYYTNATHGAYAAARATHGARGGNGCNWSNYTNATHAAYGARGAYWSN  1200
           L  I  Q  D  L  I  D  K  I  E  G  A  S  L  I  N  E  D  S
Dt         L  I  Q  D  L  I  D  K  I  E  G  A  S  L  I  N  E  D  S

Qy 1162  TTAACGATACAAGATTTAATTGATAAAATTGGAGGTGCATCGTTCATTAATGAAGATCTT  1221
           L  T  I  Q  D  L  I  D  K  I  G  G  A  S  F  I  N  E  D  L
Qt         L  T  I  Q  D  L  I  D  K  I  G  G  A  S  F  I  N  E  D  L

Db 1201  ATHAAYGTNYCARCARTTYGAYAAYGCNGTNGTNYNTTYAARGGNCNGCNGGNGAYATH  1260
           I  N  V  L  E  Q  F  D  N  A  V  V  L  F  K  G  A  G  D  I
Dt         I  N  V  L  E  Q  F  D  N  A  V  V  L  F  K  G  A  G  D  I

Qy 1222  ATTAATGTATTAGAACAATTTGATAATGCTGTGTTTTATTATGGGTGCAGGTGATATT  1281
           I  N  V  L  E  Q  F  D  N  A  V  V  L  F  M  G  A  G  D  I
Qt         I  N  V  L  E  Q  F  D  N  A  V  V  L  F  M  G  A  G  D  I

Db 1261  CARAARYNCARAAYGCNTAYYTNGAYAAARYNNGGNATGAARAAYGCNTTY  1311
           Q  K  L  Q  N  A  Y  L  D  K  L  G  M  K  N  A  F
Dt         Q  K  L  Q  N  A  Y  L  D  K  L  G  M  K  N  A  F

Qy 1282  CAAAAATTACAAAATGCATATTTAGATAAATTAGGCATGAAAAATGCCGTTT  1332
           Q  K  L  Q  N  A  Y  L  D  K  L  G  M  K  N  A  F
Qt         Q  K  L  Q  N  A  Y  L  D  K  L  G  M  K  N  A  F
```

RESULT 2
ID   MURC_BACSU   STANDARD;   PRT;   432 AA.
AC   P40778;
DT   01-FEB-1995 (REL. 31, CREATED)
DT   15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   UDP-N-ACETYLMURAMATE--ALANINE LIGASE (EC 6.3.2.8) (UDP-N-
DE   ACETYLMURANOYL-L-ALANINE SYNTHETASE).
GN   MURC.
OS   BACILLUS SUBTILIS.
OC   BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC   BACILLUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 98048467.

```
Qy   22  ATGACACACTATCATTTGTCGGAATTAAAGGTTCTGGCAATGGSGGSSLAQI  81
          M  T  H  Y  H  F  V  G  I  K  G  S  G  M  S  S  L  A  Q  I
Qt        M  T  H  Y  H  F  V  G  I  K  G  S  G  M  S  S  L  A  Q  I

Db   61  ATGCAYGAYGTNGGNCAYGARGGNWSNGAYATHGARAAYTAYGTNTTYACNGAR  120
          M  H  D  L  G  H  E  V  Q  G  S  D  I  E  N  Y  V  F  T  E
Dt        M  H  D  L  G  H  E  V  Q  G  S  D  I  E  N  Y  V  F  T  E

Qy   82  ATGCATGATTTAGGACATGAAGTTCAAGGATCGGATATTGAGAACTACGTATTTACAGAA  141
          M  H  D  L  G  H  E  V  Q  G  S  D  I  E  N  Y  V  F  T  E
Qt        M  H  D  L  G  H  E  V  Q  G  S  D  I  E  N  Y  V  F  T  E

Db  121  GTNGCNYTNCGNAAYGGNAAYGCNAAYATHAYHYNCCNTTYGAYGCNAAYAYAYATHAARGAR  180
          V  A  L  R  N  K  G  I  K  I  L  P  F  D  A  N  N  I  K  E
Dt        V  A  L  R  N  K  G  I  K  I  L  P  F  D  A  N  N  I  K  E

Qy  142  GTTGCTCTTAGAAATAACATTACCATTTGGTGCTAATAACATAAAAGAA  201
          V  A  L  R  N  K  G  I  K  I  L  P  F  G  A  N  N  I  K  E
Qt        V  A  L  R  N  K  G  I  K  I  L  P  F  G  A  N  N  I  K  E

Db  181  GAYATGGTNGTNATHCARGGNAAYGCNTTYGCNWSNWSNCAYGARGARAARGCNMGNGCN  240
          D  M  V  V  I  Q  G  N  A  F  A  S  S  H  E  E  K  A  R  A
Dt        D  M  V  V  I  Q  G  N  A  F  A  S  S  H  E  E  K  A  R  A

Qy  202  GATATGGTTATACAAGGTAATGCAATTCGCCGAGTAGCCATGAAGAAAATAGTACGTGCA  261
          D  M  V  V  I  Q  G  N  A  F  A  S  S  H  E  E  I  V  R  A
Qt        D  M  V  V  I  Q  G  N  A  F  A  S  S  H  E  E  I  V  R  A

Db  241  CAYCARATGAARYTNGAYGTNGTNWSNTAYAAYGAYTTYYTNGGNCARATHATHGAYCAR  300
          H  Q  M  K  L  D  V  V  S  Y  N  D  F  L  G  Q  I  I  D  Q
Dt        H  Q  M  K  L  D  V  V  S  Y  N  D  F  L  G  Q  I  I  D  Q

Qy  262  CATCAATTGAAATTAGATGTTGTAAGTTATAATGATTTTTTAGGACAGATTATTGATCAA  321
          H  Q  L  K  L  D  V  V  S  Y  N  D  F  L  G  Q  I  I  D  Q
Qt        H  Q  L  K  L  D  V  V  S  Y  N  D  F  L  G  Q  I  I  D  Q

Db  301  TAYACNWSNGTNGCNGTNACNGGNGCNCAYGGNAARACNWSNACNACNGGNYTNYTNWSN  360
          Y  T  S  V  A  V  T  G  A  H  G  K  T  S  T  T  G  L  L  S
Dt        Y  T  S  V  A  V  T  G  A  H  G  K  T  S  T  T  G  L  L  S

Qy  322  TATACTTCAGTAGCTGTAACTGGTGCACATGGTAAAACTTCTCAAACCGGTTTATTATCA  381
          Y  T  S  V  A  V  T  G  A  H  G  K  T  S  T  T  G  L  L  S
Qt        Y  T  S  V  A  V  T  G  A  H  G  K  T  S  T  T  G  L  L  S

Db  361  CAYGTNATGAAYGGNGAYAARAARACNWSNTTYYTNATHGAYGGNGAYGGNAACNGGNATGGGN  420
          H  V  M  N  G  D  K  K  T  S  F  L  I  I  G  D  G  T  G  M  G
Dt        H  V  M  N  G  D  K  K  T  S  F  L  I  I  G  D  G  T  G  M  G

Qy  382  CATGTTATGAATGGTGATAAAAAGACTTCATTTTTAATTATTGGTGATGGCACAGGTATGGGA  441
          H  V  M  N  G  D  K  K  T  S  F  L  I  I  G  D  G  T  G  M  G
Qt        H  V  M  N  G  D  K  K  T  S  F  L  I  I  G  D  G  T  G  M  G

Db  421  YTNCCNGARWSNGAYTAYTTYGCNTTYGARGCNTGYGARTAYMGNMGNCAYTTYYTNWSN  480
          L  P  E  S  D  Y  F  A  F  E  A  C  E  Y  R  R  H  F  L  S
Dt        L  P  E  S  D  Y  F  A  F  E  A  C  E  Y  R  R  H  F  L  S

Qy  442  TTGCCTGAAAGTGATTATTTCGCTTTTGAGGCATGTGAATATAGACGTCACTTTTTAAGT  501
          L  P  E  S  D  Y  F  A  F  E  A  C  E  Y  R  R  H  F  L  S
Qt        L  P  E  S  D  Y  F  A  F  E  A  C  E  Y  R  R  H  F  L  S

Db  481  TAYAARCCNGAYTAYGCNATHATGACNAAYATHGAYTTYGAYCAYCCNGAYTAYTTYAAR  540
          Y  K  P  D  Y  A  I  M  T  N  I  D  F  D  H  P  D  Y  F  K
Dt        Y  K  P  D  Y  A  I  M  T  N  I  D  F  D  H  P  D  Y  F  K

Qy  502  TATAAACCTGATTACGCAATTATGACAAATATTGATTTCGATCATCCTGATTATTTCAAA  561
          Y  K  P  D  Y  A  I  M  T  N  I  D  F  D  H  P  D  Y  F  K
Qt        Y  K  P  D  Y  A  I  M  T  N  I  D  F  D  H  P  D  Y  F  K

Db  541  GAYATHAAYGAYGTNTTYGAYGCNTTYCARGARATGGCNCAYAAYGTNAARAARGGNATH  600
          D  I  N  D  V  F  D  A  F  Q  E  M  A  H  N  V  K  K  G  I
Dt        D  I  N  D  V  F  D  A  F  Q  E  M  A  H  N  V  K  K  G  I

Qy  562  GATATTAATGATGTTTTTGATGCATTCCAAGAAATGGCACATAATGTTAAAAAAGGTATT  621
          D  I  N  D  V  F  D  A  F  Q  E  M  A  H  N  V  K  K  G  I
Qt        D  I  N  D  V  F  D  A  F  Q  E  M  A  H  N  V  K  K  G  I

Db  601  ATHGCNTGGGGNGAYGAYGARCAYYTNMGNAARATHGARGCNYAYGCNGAYGTNCCNATHTAYTAY  660
          I  A  W  G  D  D  E  H  L  R  K  I  E  A  D  V  P  I  Y  Y
Dt        I  A  W  G  D  D  E  H  L  R  K  I  E  A  D  V  P  I  Y  Y

Qy  622  ATTGCTTGGGGTGATGATGAACATCTACGTAAAATTGAAGCAGATGTTCCAATTTATTAC  681
          I  A  W  G  D  D  E  H  L  R  K  I  E  A  D  V  P  I  Y  Y
Qt        I  A  W  G  D  D  E  H  L  R  K  I  E  A  D  V  P  I  Y  Y

Db  661  TAYGGNAAYTAYAARGAYWSNGAYGAYATHTAYGCNCARAAYATHCARATHACNGAYAARGGN  720
          Y  G  F  K  D  S  D  D  I  Y  A  Q  N  I  Q  I  T  D  K  G
Dt        Y  G  F  K  D  S  D  D  I  Y  A  Q  N  I  Q  I  T  D  K  G

Qy  682  TATGGATTTAAAGATTCGGATGATATTTATGCTCAAAATATTCAAATTACGGATAAAGGT  741
          Y  G  F  K  D  S  D  D  I  Y  A  Q  N  I  Q  I  T  D  K  G
Qt        Y  G  F  K  D  S  D  D  I  Y  A  Q  N  I  Q  I  T  D  K  G

Db  721  ACNGCNTTYGAYGTNTAYGTNGAYGGNGARTTYTAYGAYCAYTTYYTNWSNCCNCARTAY  780
          T  A  F  D  V  Y  V  D  G  E  F  Y  D  H  F  L  S  P  Q  Y
Dt        T  A  F  D  V  Y  V  D  G  E  F  Y  D  H  F  L  S  P  Q  Y
```

```
RA   LAPIDUS A., GALLERON N., SOROKIN A., EHRLICH S.D.;
RT   "Sequencing and functional annotation of the Bacillus subtilis genes
RT   in the 200 kb rrnB-dnaB region.";
RL   MICROBIOLOGY 143:3431-3441(1997).
RN   [2]
RP   SEQUENCE OF 85-432 FROM N.A.
RC   STRAIN=168 / MARBURG;
RX   MEDLINE; 96310371.
RA   VARON D., BRODY M.S., PRICE C.W.;
RT   "Bacillus subtilis operon under the dual control of the general
RT   stress transcription factor sigma B and the sporulation transcription
RT   factor sigma H.";
RL   MOL. MICROBIOL. 20:339-350(1996).
CC   -!- FUNCTION: CELL WALL FORMATION.
CC   -!- CATALYTIC ACTIVITY: ATP + UDP-N-ACETYLMURAMOYL + L-ALANINE =
CC       ADP + ORTHOPHOSPHATE + UDP-N-ACETYLMURAMOYL-L-ALANINE.
CC   -!- PATHWAY: PEPTIDOGLYCAN BIOSYNTHESIS.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).
CC   -!- SIMILARITY: BELONGS TO THE MURCDEF FAMILY.
CC   -------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------------
DR   EMBL; AF008220; G2293216; -
DR   EMBL; L31845; G556014; -
SUBTILIST; BG10973; MURC.
KW   PEPTIDOGLYCAN SYNTHESIS; CELL WALL; CELL DIVISION; LIGASE;
KW   ATP-BINDING.
FT   NP_BIND     108     114     ATP (POTENTIAL)
SQ   SEQUENCE    432 AA;   48364 MW;   163166CB CRC32;

Query Match             43.9%;
Best Local Similarity   47.8%;    Score 2964;  DB 1;  Length 432;
                                  Pred. No. 0.00e+00;
Matches   605;   Conservative  248;   Mismatches 414;   Indels  0;   Gaps  0;
```

```
Db    1  M  T  V  Y  H  F  V  G  I  K  G  T  G  M  S  P  L  A  Q  I
Dt    1  ATGACNGTNTAYCAYTTYGTNGGNATHAARGGNACNGGNATGWSNCCNYTNGCNCARATH   60
         ||||   ::|| |||:||:||: |||  ||  ||   || || :||  :|  :||:|:|
Qy   22  ATGACACACTATCATTTGTCGGAATTAAAGTTCTCGCAGTTCATTAGCACAAATC       81
Qt       M  T  H  Y  H  F  V  G  I  K  G  S  G  M  S  S  L  A  Q  I

Db   61  L  H  D  N  G  Y  T  V  Q  G  S  D  I  E  K  F  I  F  T  Q
Dt       YTNCAYGAYAAYGGNTAYACNGTNCARGGNWSNGAYATHGARAARTTYATHTTYACNCAR  120
         ::|   |   ||::  ::|  |  || ||   :| || || :| || || ||  :||:|
Qy   82  ATGCATGATTTAGGACATGAAGTTCAAGGATCGGATATTGAAAAGTTTATTTTTACAGAA 141
Qt       M  H  D  L  G  H  E  V  Q  G  S  D  I  E  N  Y  V  F  T  E

Db  121  T  A  L  E  K  R  N  I  T  I  L  P  F  S  A  E  N  I  K  P
Dt       ACNGCNYTNGARAARMGNAAYATHACNATHYTNCCNTTYWSNGCNGARAAYATHAARCCN  180
         ::|| |::|  || |: ||:|| || || ||:|| ||  || || || || || || :::
Qy  142  GTTGCTCTTAGAAAATAGGGGATAAAATATTACCATTGGTGCTAATAACATAAAAGAA   201
Qt       V  A  L  R  N  K  G  I  K  I  L  P  F  G  A  N  N  I  K  E

Db  181  G  M  T  V  I  A  G  N  A  F  P  D  T  H  P  E  I  E  K  A
Dt       GGNATGACNGTNATHGCNGGNAAYGCNTTYCCNGAYACNCAYCCNGARATHGARAARGCN  240
         || ||:|| || || || || ||:||:||  ||  || || ||:|| |: || || |
Qy  202  GATATGGTAGTTATCACAAGGATAATGCATTCGCCGAGTACGTATTTTAGGAACTACCTGCA 261
Qt       D  M  V  V  I  Q  G  N  A  F  A  S  S  H  E  E  I  V  R  A

Db  241  M  S  E  G  I  P  V  I  R  Y  H  K  F  L  G  D  Y  M  K  K
Dt       ATGWSNGARGGNATHCCNGTNATHMGNTAYCAYAARTTYYTNGGNGAYTAYATGAARAAR  300
         || :| || || || ||  ||:|| ||:|| ||:|| ||  || ||:|| ||:|| || |
Qy  262  CATCAATTGAAATTAGATGTTGTAAGTTATAATGATTTTTTAGGACAGATTATTGATCAA  321
Qt       H  Q  L  K  L  D  V  V  S  Y  N  D  F  L  G  Q  I  I  D  Q

Db  301  F  T  S  V  A  V  T  G  A  H  G  K  T  S  T  T  G  L  L  A
Dt       TTYACNWSNGTNGCNGTNACNGGNGCNCAYGGNAARACNWSNACNACNGGNYTNTTNGCN  360
         :|:|| ::|:|| || || ||  || || ||  || || :| || || || ::|::|:|
Qy  301  TTTACTTCAGTAGCTGTACTGGTGCACATGGTGAAACTTCTACAACAGTTTATTATCA   381
Qt       Y  T  S  V  A  V  T  G  A  H  G  K  T  S  T  T  G  L  L  S

Db  361  H  V  I  Q  N  A  K  P  T  S  F  L  I  G  D  T  G  Q  G
Dt       CAYGTNATHCARAAYGCNAARCCNACNWSNTTYYTNATHGGNGAYGGNCANGGNCARGGN  420
         || || ||:|| || || || ||:|| :| ||:::|| || ||:||  || || :
Qy  382  CATGTTATGAATGGTGATAAAAAGACTTCATTTTTAATTGGTGATGGCACAGGTATGGGA  441
Qt       H  V  M  N  G  D  K  K  T  S  F  L  I  G  D  G  T  G  M  G

Db  421  N  E  N  S  E  Y  F  V  F  E  A  C  E  Y  R  R  H  F  L  S
Dt       AAYGARAAYWSNGARTAYTTYGTNTTYGARGCNTGYGARTAYMGNMGNCAYTTYYTNWSN  480
         || || || ||:|| || ||:|| || || || :| || ||:|: |: ||:||:::| ||
Qy  442  TTGCCTGAAAGTGATTATTTCGCTTTGAGGCCATGTGAATATAGACGTCACTTTTAAGT   501
Qt       L  P  E  S  D  Y  F  A  F  E  A  C  E  Y  R  R  H  F  L  S

Db  481  Y  Q  P  D  Y  A  I  M  T  N  I  D  F  D  H  P  D  Y  F  S
Dt       TAYCARCCNGAYTAYGCNATHATGACNAAYATHGAYTTYGAYCAYCCNGAYTAYTTYWSN  540
         || |: || ||:|| || || ||:|| || ||:|| ||:|| || ||:||:|| ||  :
Qy  481  TAYAAACCTGATTACGCAATTATGACAAATATTGATTTCGATCATCCTGATTATTTCAAA   5
Qt       Y  K  P  D  Y  A  I  M  T  N  I  D  F  D  H  P  D  Y  F  K

Db  541  S  I  D  D  V  F  D  A  F  Q  E  M  A  L  Q  V  N  K  G  I
Dt       WSNATHGAYGAYGTNTTYGAYGCNTTYCARGARATGGCNYTNCARGTNAAYAARGGNATH  600
         :||::|||:||:|| ||:||:|| || || || || || :|:||:|:  || || || |
Qy  541  GATATTAATGATGTTTTTGATGCATTCCAAGAAATGGCACATAATGTTAAAAAGGTATT   621
Qt       D  I  N  D  V  F  D  A  F  Q  E  M  A  H  N  V  K  K  G  I

Db  601  I  A  C  G  D  D  E  H  L  P  K  I  H  A  N  V  P  V  V  Y
Dt       ATHGCNTGYGGNGAYGAYGARCAYYTNCCNAARATHCAYGCNAAYGTNCCNGTNGTNTAY  660
         || || :| || ||:|| ||:||:::| |: || ||::| :| ||:|| || |   |  |
Qy  601  ATTGCTTGGGGTGATGATGAACATCTACGTAAAATTGAAGCAGATGTTCCAATTTATTAC   681
Qt       I  A  W  G  D  D  E  H  L  R  K  I  E  A  D  V  P  I  Y  Y

Db  661  Y  G  T  G  E  E  N  D  F  Q  A  R  N  I  V  K  S  T  E  G
Dt       TAYGGNACNGGNGARGARAAYGAYTTYCARGCNMGNAAYATHGTNAARWSNACNGARGGN  720
         || || ::| |  |  :| || ||  |    |  :: :: || :| || :: || :| |
Qy  682  TATGGATTTAAAGATTCGGATGACATTTATGCTCAAAATATTCAAATTACGGATAAAGGT  741
Qt       Y  G  F  K  D  S  D  D  I  Y  A  Q  N  I  Q  I  T  D  K  G

Db  721  T  T  F  D  V  F  V  R  N  T  F  Y  D  T  F  Y  I  P  A  Y
Dt       ACNACNTTYGAYGTNTTYGTNMGNAAYACNTTYTAYGAYACNTTYTAYATHCCNGCNTAY  780
         || |  || || ||:|:  :  || || || || || || || |   ::|||:|| ||:|
Qy  742  ACTGCTTTTGATGTGTATGTGGTGGATGGTGAGTTTTATGATCACTTCCTGTCTCCACAATAT 801
Qt       T  A  F  D  V  Y  V  V  D  G  E  F  Y  D  H  F  L  S  P  Q  Y

Db  781  G  H  H  N  V  L  N  S  L  A  V  I  A  L  C  H  Y  E  E  I
Dt       GGNCAYCAYAAYGTNYTNAAYWSNYTNGCNGTNATHGCNYTNTGYCAYTAYGARGARATH  840
         || :| ||  | :|:::|| ::| ::|| || || || ::| |::| ||:|| || || :
Qy  802  GGTGACCATACAGTTTTAAATGCATTAGCTGTAATTGCGATTAGTTATTTAGAGAAGCTA   861
Qt       G  D  H  T  V  L  N  A  L  A  V  I  A  I  S  Y  L  E  K  L

Db  841  D  S  S  I  I  K  H  A  L  K  S  F  G  G  V  K  R  F  N
Dt       GAYWSNWSNATHATHAARCAYGCNYTNAARWSNTTYGGNGGNGTNAARMGNTTYAAY     900
         || :| ::||:|| || || || || || |  ||:|| || || || |: ||:|| |
Qy  862  GATGTTACAAATATTAAAGAAGCATTAGAAACGTTTGGTGGTGTTAAACGTCGTTTCAAT   921
Qt       D  V  T  N  I  K  E  A  L  E  T  F  G  G  V  K  R  R  F  N

Db  901  E  K  Q  L  G  D  Q  V  L  I  D  D  Y  A  H  H  P  T  E  I
Dt       GARAARCARYTNGGNGAYCARGTNYTNATHGAYGAYTAYGCNCAYCACNGARATHGCN     960
         || || || :| || ||:|| || :| |: ||:||:|| || ||:|| || || :| |
Qy  922  GAAACTACAATTGCAAATCAAGTTATTGTAGATGATTATGCACACCATCCAAGAGAAATT  981
Qt       E  T  T  I  A  N  Q  V  I  V  D  D  Y  A  H  H  P  R  E  I

Db  961  K  V  T  I  E  A  A  R  Q  K  Y  P  D  R  E  I  V  A  V  F
Dt       AARGTNACNATHGARGCNGCNMGNCARAARTAYCCNGAYMGNGARATHGTNGCNGTNTTY  1020
         || || |  ::||: |  || |: || || || || || |: |: || || || || ||
Qy  982  AGTGCTACAATTGACACAGCACGAAAGAAATATCCACATAAAGAAGTTGTTGCAGTATTT  1041
Qt       S  A  T  I  D  T  A  R  K  K  Y  P  H  K  E  V  V  A  V  F

Db 1021  Q  P  H  T  F  T  R  T  Q  Q  F  L  D  E  F  A  E  S  L  S
Dt       CARCCNCAYACNTTYACNMGNACNCARCARTTYYTNGAYGARTTYGCNGARWSNYTNWSN  1080
         || |  || || ||:|  |: |  || || || ::| || || ||  || :| ::|:| ||
Qy 1042  CAACCACACACTTTCTCTAGAACACAGCATTTTAAATGAATTGCAGAGAATTTATGT  1101
Qt
```

Best Local Similarity 40.5%;   Pred. No. 8.29e-24;
Matches 77;   Conservative 38;   Mismatches 75;   Indels 0;   Gaps 0;

```
                         V   L   I   D   D   Y   A   H   P   V   E   L   D   A   A   I   R   S   V
Db   973 GTNYTNATHGAYGAYTAYGCNCAYCAYCCNGTNGARYTNGAYGCNGCNATHMGNWSNGTN 1032
            ||| :|:||:||||:|:|| ||:||:|| ||:||   :||  :|| ||  ||:| :
Dt   943 GTTATTGTAGATGATTATGCACACCCTACACAGAGAAATTAGTGCTACAATTGACACAGCA 1002
                         V   I   V   D   D   Y   A   H   H   P   R   E   I   S   A   T   I   D   T   A
Qy
Qt

                         R   E   I   Y   S   G   K   H   I   M   G   I   F   Q   P   H   L   Y   S   R
Db  1033 MGNGARATHTAYWSNGGNAARCAYATHATGGGNATHTTYCARCCNCAYYTNTAYWSNMGN 1092
            :|:||:|:  ||  :|:|||  || ||:|| |||||||||:||||| ||||||:|:||:
Dt  1003 CGAAGAAATATCCACATAAAGAAGAAGTTGTTGCAGTATTTCAACCACACACTTTCTCTAGA 1062
                         R   K   K   Y   P   H   K   E   V   V   A   V   F   Q   P   H   T   F   S   R
Qy
Qt

                         T   A   D   F   Y   Q   D   F   A   K   S   L   S   M   L   D   Q   V   V   L
Db  1093 ACNGCNGAYTTYTAYCARGAYTTYGCNAARWSNYTNWSNATGYTNGAYCARGTNGTNYTN 1152
            ||| ||  :| |||:|||||:|| ||:|  :|| :|:||| || ||:||  ||:|| ||
Dt  1063 ACACAAGCATTTTAAATGAATTTGCAGAAGTTTAGTGAAGCAGATCGTGTATTCTTA 1122
                         T   Q   A   F   L   N   E   F   A   E   S   L   C   K   A   D   R   V   F   L
Qy
Qt

                         L   D   I
Db  1153 YTNGAYATHT 1162
            : |||:|
Dt  1123 TGTGAAATTT 1132
                         C   E   I
Qy
Qt
```

RESULT 4
ID   MURC_SYNY3   STANDARD;   PRT;   505 AA.
AC   P74528;
DT   15-JUL-1998 (REL. 36, CREATED)
DT   15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   UDP-N-ACETYLMURAMATE--ALANINE LIGASE (EC 6.3.2.8) (UDP-N-
DE   ACETYLMURANOYL-L-ALANINE SYNTHETASE).
GN   MURC OR SLR1423.
OS   SYNECHOCYSTIS SP. (STRAIN PCC 6803).
OC   BACTERIA; CYANOBACTERIA; CHROOCOCCALES; SYNECHOCYSTIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97061201.
RA   KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
RA   MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,
RA   HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARUO K., OKUMURA S.,
RA   SHIMPO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M.,
RA   TABATA S.;
RT   "Sequence analysis of the genome of the unicellular cyanobacterium
RT   Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT   entire genome and assignment of potential protein-coding regions.";
RL   DNA RES. 3:109-136(1996).
CC   -!- FUNCTION: CELL WALL FORMATION (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: ATP + UDP-N-ACETYLMURAMOYL + L-ALANINE =
CC   ADP + ORTHOPHOSPHATE + UDP-N-ACETYLMURAMOYL-L-ALANINE.
CC   -!- PATHWAY: PEPTIDOGLYCAN BIOSYNTHESIS.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).
CC   -!- SIMILARITY: BELONGS TO THE MURCDEF FAMILY.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; D90916; G1653723; -.
KW   PEPTIDOGLYCAN SYNTHESIS; CELL WALL; CELL DIVISION; LIGASE;
KW   ATP-BINDING.
FT   NP_BIND   164   170   ATP (POTENTIAL).
SQ   SEQUENCE   505 AA;   54513 MW;   B3D7FC7C CRC32;

Query Match   4.6%;   Score 309;   DB 1;   Length 505;

```
                     Q   P   H   T   F   S   R   T   Q   A   F   L   N   E   F   A   E   S   L   C
Qt
Db  1081 GGNGCNGAYTGYGTNAYTNTGAYATHTTYGGNWSNGCNMGNGARAAYGCNGGNAAR 1140
            ||:||  :|:|| :||||||||||||:|| :: :| ||||||:||||| |||
Dt  1102 AAAGCAGATCGTGTATTCTTATGTGAAATTTTGGCTCAATTAGAGAAAATTCTGGCGCA 1161
                     K   A   D   R   V   F   L   C   E   I   F   G   S   I   R   E   N   S   G   A
Qy
Qt

                     L   T   I   I   G   D   L   Q   G   K   I   H   N   A   K   L   I   E   D   D
Qt
Db  1141 YTNACNATHGGNGAYTNCARGGNAARATHCAYAAYGCNAARYTNATHGARGARCAYGAY 1200
            :|:||||:|||||  :| :||||||||| |:|: ||:|||| :| |:|:||:||:||:
Dt  1162 TTACGATACAAGATTAATTGATAAAATTGGAGGTGCATCGTTCATTAATGAAGATCTT 1221
                     L   T   I   Q   D   L   I   D   K   I   G   G   A   S   I   N   E   D   L
Qy
Qt

                     T   S   V   L   K   A   H   D   K   A   V   L   I   F   M   G   A   G   D   I
Qt
Db  1201 ACNWSNGTNYTNAARGCNCAYGAYAARGCNGTNYTNATHTTYATGGGNGCNGGNGAYATH 1260
            ||  :| ||:||||:||:|||||||||||| |||:| ||:|||||||| ||||:||:|||
Dt  1222 ATTAATGTATTAGAACAAATTTGATAATGCTGTTTATTAATGGTGCCAGGTGATATT 1281
                     I   N   V   L   E   Q   F   D   N   A   V   V   L   F   M   G   A   G   D   I
Qy
Qt

                     Q   K
Qt
Db  1261 CARAART 1267
            ||:||||
Dt  1282 CAAAAAT 1288
                     Q   K
Qy
Qt
```

RESULT 3
ID   MURC_PORGI   STANDARD;   PRT;   433 AA.
AC   Q51831;
DT   15-JUL-1998 (REL. 36, CREATED)
DT   15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   UDP-N-ACETYLMURAMATE--ALANINE LIGASE (EC 6.3.2.8) (UDP-N-
DE   ACETYLMURANOYL-L-ALANINE SYNTHETASE).
GN   MURC.
OS   PORPHYROMONAS GINGIVALIS (BACTEROIDES GINGIVALIS).
OC   BACTERIA; CYTOPHAGALES; BACTEROIDACEAE; PORPHYROMONAS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=381;
RX   MEDLINE; 96118684.
RA   ANSAI T., YAMASHITA Y., AWANO S., SHIBATA Y., WACHI M., NAGAI K.,
RA   TAKEHARA T.;
RT   "A murC gene in Porphyromonas gingivalis 381.";
RL   MICROBIOLOGY 141:2047-2052(1995).
RN   [2]
RP   REVISIONS.
RC   STRAIN=381;
RA   ANSAI T.;
RL   SUBMITTED (JAN-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- FUNCTION: CELL WALL FORMATION (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: ATP + UDP-N-ACETYLMURAMOYL + L-ALANINE =
CC   ADP + ORTHOPHOSPHATE + UDP-N-ACETYLMURAMOYL-L-ALANINE.
CC   -!- PATHWAY: PEPTIDOGLYCAN BIOSYNTHESIS.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).
CC   -!- SIMILARITY: BELONGS TO THE MURCDEF FAMILY.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; D84504; D1025270; -.
KW   PEPTIDOGLYCAN SYNTHESIS; CELL WALL; CELL DIVISION; LIGASE;
KW   ATP-BINDING.
FT   NP_BIND   114   120   ATP (POTENTIAL).
SQ   SEQUENCE   433 AA;   48309 MW;   38C4714B CRC32;

Query Match   4.6%;   Score 310;   DB 1;   Length 433;
```

```
 __   __  ___   _   ___    __ 
|  \/  ||  _ \ / \ |  _ \  / _|
| |\/| || |_) / _ \| |_) ||  _|
| |  | ||  __/ ___ \  __/ | |_ 
|_|  |_||_|  /_/   \_\_|   \___| (TM)
```

****

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

****

MPsrch_pp   protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Nov 24 00:36:34 1999;   MasPar time 19.53 Seconds
                                         896.452 Million cell updates/sec
Tabular output not generated.

Title: >US-09-103-287-2
Description: (1-437) from US09103287.pep
Perfect Score: 3121
Sequence: 1 MTHYHFVGIKGSGMSSLAQI.........GDIQKLQNAYLDKLGMKNAF 437

Scoring table: PAM 150
                   Gap 11

Searched: 122810 seqs, 40068593 residues

Statistics: Mean 50.576; Variance 122.235; scale 0.414

Post-processing: Minimum Match 0%
                   Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                        SUMMARIES

Database: pir60
                   1:pir1 2:pir2 3:pir3 4:pir4

         %
Result   Query
No. Score Match Length DB ID   Description   Pred. No.
----
  1 1975 63.3   432 2 C69662   UDP-N-acetylmuramate-   0.00e+00
  2   620 19.9   454 2 B70418   UDP-N-acetylmuramate-   2.69e-77
  3   602 19.3   481 2 C71338   probable UDP-N-acetyl   2.58e-74
  4   584 18.7   495 2 C71679   probable UDP-N-acetyl   2.42e-71
  5   581 18.6   468 2 H70201   UDP-n-acetylmuramate-   7.56e-71
  6   567 18.2   803 2 A71475   probable muramate-Ala   1.53e-68
  7   529 16.9   475 2 E64185   UDP-N-acetylmuramate-   1.60e-62
  8   495 15.9   491 1 CEECAM   UDP-N-acetylmuramoyla   8.97e-57
  9   470 15.1   505 2 S76722   UDP-N-acetylmuramoyla   9.98e-53
 10   440 14.1   453 2 B64002   hypothetical 48.5K pr   6.64e-48
 11   424 13.6   457 2 S56459   hypothetical 48.5K pr   2.39e-45
 12   422 13.5   449 2 G64597   UDP-N-acetylmuramate-   4.98e-45
 13   409 13.1   449 2 B71917   udp-n-acetylmuramate-   5.81e-43
 14   349 11.2   494 2 D70579   probable murC protein   1.52e-33
 15   180   5.8   494 2 B47691   UDP-N-acetylmuramoyla   7.76e-09
 16   177   5.7   457 2 I64184   UDP-N-acetylmuramoyla   1.93e-08
 17   160   5.1   445 2 A71699   UDP-n-acetylmuramoyla   3.05e-06
 18   158   5.1   505 2 S75968   UDP-N-acetylmuramoyla   5.45e-06
 19   154   4.9   682 2 A69170   UDP-N-acetylmuramyl t   1.73e-05
 20   147   4.7   450 2 B69198   UDP-N-acetylmuramyl t   1.27e-04
 21   142   4.5   508 2 A70125   UDP-N-acetylmuramoyla   5.11e-04
 22   128   4.1   445 2 G70371   UDP-MURNAC-pentapepti   2.23e-02
 23   128   4.1   449 2 JC6560   UDP-N-acetylmuramoyla   2.23e-02
```

```
24  128  4.1  457  2 F69662  UDP-N-acetylmuramoyla  2.23e-02
25  124  4.0  856  2 A64699  hypothetical protein   6.32e-02
26  121  3.9  464  2 H70137  UDP-N-acetylmuramoyla  1.36e-01
27  118  3.8  451  2 D47691  UDP-N-acetylmuramoyla  2.90e-01
28  119  3.8  482  2 E70450  UDP-MurNac-tripeptide  2.26e-01
29  120  3.8  598  2 D71127  hypothetical protein   1.75e-01
30  116  3.7  437  2 B64185  UDP-N-acetylmuramoyla  4.77e-01
31  111  3.6  121  2 S18067  thyroid hormone recep  1.62e+00
32  111  3.6  320  2 S33172  ornithine cyclodeamin  9.96e-01
33  111  3.6  432  2 H64383  Na+ transporter - Met  1.62e+00
34  111  3.6  493  2 D64612  UDP-MurNac-pentapepti  1.62e+00
35  108  3.5  143  2 E69065  molybdenum cofactor b  3.30e+00
36  109  3.5  452  2 F64730  UDP-N-acetylmuramoyla  2.61e+00
37  109  3.5  585  2 C70330  conserved hypothetica  2.61e+00
38  110  3.5 1558  2 B71603  RESA-H3 antigen PFB09  2.05e+00
39  110  3.5 1670  2 S71551  DNA-directed DNA poly  2.05e+00
40  105  3.4  227  2 H64336  formate dehydrogenase  6.65e+00
41  106  3.4  374  2 S75646  hypothetical protein   5.28e+00
42  106  3.4  416  2 E69599  competence-damage ind  5.28e+00
43  106  3.4  447  2 D71812  udp-n-acetylmuramyl-t  5.28e+00
44  106  3.4  447  2 F64706  UDP-MurNac-tripeptide  5.28e+00
45  105  3.4  758  2 S37855  hypothetical protein   6.65e+00
```

```
                 ALIGNMENTS

RESULT      1
ENTRY       C69662      #type complete
TITLE       UDP-N-acetylmuramate-alanine ligase (EC 6.3.2.8) murC -
            Bacillus subtilis
ORGANISM    #formal_name Bacillus subtilis
DATE        05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
            17-Mar-1999
ACCESSIONS  C69662; S71002
REFERENCE   A69580
 #authors   Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
            Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
            Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
            A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, N.M.;
            Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
            Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;
            Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.;
            Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.;
            Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
            M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim,
            S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.;
            Guiseppi, G.; Guy, B.J.; Haga, K.; Haiech, J.; Harwood,
            C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
            Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
            Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi,
            Y.; Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.;
            Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
            Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
            Mauel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
            M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
            M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro,
            V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott,
            A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;
            Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
            Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.;
            Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
            Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo,
            B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;
            Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;
            Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
            Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
            Wambutt, R.; Wedler, E.; Wedler, H.; Weitenegger, T.;
            Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
            K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;
            Yoshikawa, H.; Danchin, A.
 #journal   Nature (1997) 390:249-256
 #title     The complete genome sequence of the Gram-positive bacterium
            Bacillus subtilis.
```

#cross-references MUID:98044033
#accession   C69662
##status   preliminary; nucleic acid sequence not shown;
             translation not shown
##molecule_type DNA
##residues   1-432 ##label KUN
##cross-references GB:299119; GB:AL009126; NID:g2635411; PID:e1185852;
             PID:g2635463
##experimental_source strain 168
REFERENCE   S71000
#authors   Varon, D.; Brody, M.S.; Price, C.W.
#journal   Mol. Microbiol. (1996) 20:339-350
#title   Bacillus subtilis operon under the dual control of the
             general stress transcription factor sigma(B) and the
             sporulation transcription factor sigma(H).
#cross-references MUID:96310371
#accession   S71002
##molecule_type DNA
##residues   85-432 ##label VAR
##cross-references EMBL:L31845; NID:g556013; PID:g556014
##experimental_source strain 168, substrain Marburg
GENETICS
#gene   murC
FUNCTION
#pathway   peptidoglycan biosynthesis
KEYWORDS   ATP binding; cell division; cell wall; ligase; peptidoglycan
             biosynthesis
SUMMARY   #length 432 #molecular-weight 48364 #checksum 6127

Query Match   63.3%;   Score 1975; DB 2; Length 432;
Best Local Similarity 63.9%;   Pred. No. 0.00e+00;
Matches 273; Conservative 60; Mismatches 94; Indels 0; Gaps 0;

Db   1 MTVYHFVGIKGTGMSPLAQILHDNGYTVQGSDIEKFIFTQTALEKRNITILPFSAENIKP 60
          ||:||||||||:||||:|||:|:|:|:|||||||:::|:|:||:|:|||||||::||||
Qy   1 MTHYHFVGIKGSGMSSLAQIMHDLGHEVQGSDIENYFTEVALRNKGIKILPFGANNIKE 60

Db   61 GMTVIAGNAFPDTHPEIEKAMSEGIPVIRYHKFLGDYMKKFTSVAVTGAHGKTSTTGLLA 120
          :| ||:|||| ::||||::| : |:::||::||:||::::||||||||||||||||||:
Qy   61 DMVVIQGNAFASSHEEIVRAHQLKLDVVSYNDFLGQIIDQYTSVAVTGAHGKTSTTGLLS 120

Db 121 HVIQNAKPTSFLIGDTGQGNENSEYFVFEACEYRRHFLSYQPDYAIMTNIDFDHPDYFS 180
          ||:: :| |||||||||:| : ||||:||||||||||||||:||||||||||||||||:
Qy 121 HVMNGDKKTSFLIGDTGMGLPESDYFAFEACEYRRHFLSYKPDYAIMTNIDFDHPDYFK 180

Db 181 SIDDVFDAFQEMALQVNKGIIACGDDEHLPKIHANVPVVYGTGEENDFQARNIVKSTEG 240
          .|:||||||||||::|:|||||:|||||||:|.|:||:::| :.:|||:|||||:| :|
Qy 181 DINDVFDAFQEMAHNVKKGIIAWGDDEHLRKIEADVPIYYGFKDSDDIYAQNIQITDKG 240

Db 241 TTFDVFVRNTFYDTFYIPAYGHHNVLNSLAVIALCHYEEIDSSIIKHALKSFGGVKRRFN 300
          |:|||:| ::|| | |||||:|:|||:|||||| ::|:|:| ::|:|||:|||||||||
Qy 241 TAFDVYVDGEFYDHFLSPQYGDHTVLNALAVIAISYLEKLDVTNIKEALETFGGVKRRFN 300

Db 301 EKQLGDQVLIDDYAHHPTEIKVTIEAARQKYPDREIVAVFQPHTFTRTQQFLDEFAESLS 360
          | : :||::||||||||:| :||:|||:|||:::::||||||||:|||:||:|||||||
Qy 301 ETTIANQVIVDDYAHHPREISATIDTARKKYPHKEVVAVFQPHTFSRTQAFLNEFAESLC 360

Db 361 GADCVVLCDIFGSIRENSGALTIGDLQGKIHNAKLIEEDDTSVLKAHDKAVLIFMGAGDI 420
          :|| |.||:|||||||||||||:||::.||:.| :|:|||:.|:: |:|||:|||||||
Qy 361 KADRVFLCEIFGSIRENSGALTIQDLIDKIGGASFINEDLINVLEQFDNAVVLFMGAGDI 420

Db 421 QKYMRAY 427
          || :|||
Qy 421 QKLQNAY 427

RESULT   2
ENTRY   B70418   #type complete
TITLE   UDP-N-acetylmuramate-alanine ligase - Aquifex aeolicus
ORGANISM #formal_name Aquifex aeolicus
DATE   08-May-1998 #sequence_revision 08-May-1998 #text_change
         21-Aug-1998
ACCESSIONS B70418

REFERENCE   A70300
#authors   Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.;
             Lenox, A.L.; Graham, D.E.; Overbeek, R.; Snead, M.A.;
             Keller, M.; Aujay, M.; Huber, R.; Feldman, R.A.; Short,
             J.M.; Olson, G.J.; Swanson, R.V.
#journal   Nature (1998) 392:353-358
#title   The complete genome of the hyperthermophilic bacterium
             Aquifex aeolicus.
#cross-references MUID:98196666
#accession   B70418
##status   preliminary; nucleic acid sequence not shown;
             translation not shown
##molecule_type DNA
##residues   1-454 ##label AQF
##cross-references GB:AE000736; NID:g2983763; PID:g2983764; GB:AE000657
##experimental_source strain VF5
GENETICS
#gene   murC
CLASSIFICATION #superfamily UDP-N-acetylmuramate--alanine ligase
SUMMARY   #length 454 #molecular-weight 50893 #checksum 2346

Query Match   19.9%;   Score 620; DB 2; Length 454;
Best Local Similarity 30.8%;   Pred. No. 2.69e-77;
Matches 139; Conservative 125; Mismatches 164; Indels 24; Gaps 20;

Db   6 IKKFHFIGIGIGMSGIAQILLEMGYKVSGSDISENKNTKL-LKQGAKIYIGHRPENLIG 64
          :||:||: |||:||:|||| :|:|:|:.|||||:: .||| ||:|||:|:. :|:|| :
Qy   1 MTHYHFVGIKGSGMSSLAQIMHDLGHEVQGSDIENYFTEVALRNKGIKI-LPFGANNIK 59

Db   65 DAQVVVYSSAVKPDNPEIQEAKRRNIPVIPRGEMLAELFKLKEGIAVSGSHGKTTTTSMI 124
          : :||: .:|. :.:.||. |:. .: : .:.:| :: .. ..|:|:|:|||::||.::
Qy   60 EDMVVIQGNAFASSHEEIVRAHQLKLDVVSYNDFLGQIIDQYTSVAVTGAHGKTSTTGLL 119

Db 125 AEILINAGLEPTVIIGGRLKRLGTNAKLGRGELLVSEADESDGSFLKLQPAVAVITNVDK 184
          :.:: ||| | .: | |.: | :. . .. | |::.|.:|:::||:|
Qy 120 SHVM-NGDKKTSFLIGD----GTGMGLPESDYFAFEACEYRRHFLSYKPDYAIMTNIDF 173

Db 185 EHLDFYENFERVKEAFEQFMNSVPFYGFAVVNLDDPTLAQLVKKSHERVITGINSPALV 244
          :|.|:::: :.|::|||:|: .: : | :.|| .| ::::. .|.::.. :. .::
Qy 174 DHPDYFKDINDVFDAFQEMAHNVKK-GI-IAWGDDEHLRK-IE-ADVPIYYGFKDSDDI 229

Db 245 RAKNLYIKEGRYEFGVEFKGKELGRIHLGIA-GIHNVYNALAATGVA-LE-LGVSFEVIK 301
          .|:|: : :. :.| :.:.|.|. :|: | : .:: :|||||::|| || |.||::::.
Qy 230 YAQNIQITDKGTAFDVYVDGEFYDHF-LSPQYGDHTVLNALAVIAISYLEKLDVT-N-IK 286

Db 302 KSLEEFRNAERRLELKGYYKNSPVYDDYGHHPTEIKAVINSLRDMYPDKNLLVFQPHRY 361
          ::| ||.. :||: :.|:.:::.:|:||.|||.|| |.:: .|:.:|.|::.||||| :
Qy 287 EALETFGGVKRRFNETTI-ANQVIVDDYAHHPREISATIDTARKKYPHKEVVAVFQPHTF 345

Db 362 SRTYYLFEDFVKVLKDIDKLIVTDIYPASENNVYGVSAEELARK-SGAVFAKDKEEVFEK 420
          ||| :.:.| .:|.. ::.:: :|:.:|.|| :|::::.|:.| ||: :|.::::::
Qy 346 SRTQAFLNEFAESLCKADRVFLCEIFGSIRENSGALTIQDLIDKIGGASFI-N-EDLIN- 402

Db 421 VREVHDEGDVLLFLGAGSISKWCEEFLKEVNL 452
          | |..|: ||:||:|||. .|: .:. .::.:
Qy 403 VLEQFDNA-VVLFMGAGDIQKLQNAYLDKLGM 433

RESULT   3
ENTRY   C71338   #type complete
TITLE   probable UDP-N-acetylmuramate--alanine ligase (murC) -
         syphilis spirochete
ORGANISM #formal_name Treponema pallidum subsp. pallidum #common_name
         syphilis spirochete
DATE   24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change
         17-Mar-1999
ACCESSIONS C71338
REFERENCE   A71250
#authors   Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.;
             Sutton, G.G.; Dodson, R.; Gwinn, M.; Hickey, E.K.; Clayton,
             R.; Ketchum, K.A.; Sodergren, E.; Hardham, J.M.; McLeod,
             M.P.; Salzberg, S.; Peterson, J.; Khalak, H.; Richardson,
             D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McDonald,

```
GN   MURC.
OS   BACILLUS SUBTILIS.
OC   BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC   BACILLUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 98048467.
RA   LAPIDUS A., GALLERON N., SOROKIN A., EHRLICH S.D.;
RT   "Sequencing and functional annotation of the Bacillus subtilis genes
RT   in the 200 kb rrnB-dnaB region.";
RL   MICROBIOLOGY 143:3431-3441(1997).
RN   [2]
RP   SEQUENCE OF 85-432 FROM N.A.
RC   STRAIN=168 / MARBURG;
RX   MEDLINE; 96310371.
RA   VARON D., BRODY M.S., PRICE C.W.;
RT   "Bacillus subtilis operon under the dual control of the general
RT   stress transcription factor sigma B and the sporulation transcription
RT   factor sigma H.";
RL   MOL. MICROBIOL. 20:339-350(1996).
CC   -!- FUNCTION: CELL WALL FORMATION.
CC   -!- CATALYTIC ACTIVITY: ATP + UDP-N-ACETYLMURAMOYL + L-ALANINE =
CC       ADP + ORTHOPHOSPHATE + UDP-N-ACETYLMURAMOYL-L-ALANINE.
CC   -!- PATHWAY: PEPTIDOGLYCAN BIOSYNTHESIS.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).
CC   -!- SIMILARITY: BELONGS TO THE MURCDEF FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use  by  non-profit institutions as long  as its content is in  no  way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF008220; G2293216; -.
DR   EMBL; L31845; G556014; -.
DR   SUBTILIST; BG10973; MURC.
KW   PEPTIDOGLYCAN SYNTHESIS; CELL WALL; CELL DIVISION; LIGASE;
KW   ATP-BINDING.
FT   NP_BIND     108    114    ATP (POTENTIAL).
SQ   SEQUENCE   432 AA;  48364 MW;  163166CB CRC32;
```

```
Query Match             40.1%;  Score 1323; DB 1;  Length 432;
Best Local Similarity   45.7%;  Pred. No. 3.53e-217;
Matches  269;  Conservative  122;  Mismatches 197;  Indels   0;  Gaps   0;
```

```
Db           D  F  Q  A  R  N  I  V  K  S  T  E  G  T  T  F  D  V  F  V
Dt   680 AYGAYTTYCARGCNMGNAAYATHGTNAARWSNACNGARGGNACNACNTTYGAYGTNTTYG 739
         |:::||:|:|||: ||||||| ::   ::  |:|:::|| | :||:|:|::||:: :|:|
Qy    15 ATGACATTTATGCTCAAATATTTCAAATATCGATAAAGGTACTGCTGTTGATGTGTATG 74
Qt       D  I  Y  A  Q  I  F  Q  I  T  D  K  G  T  A  V  D  V  Y  V
```

```
Db           R  N  T  F  Y  D  T  F  Y  I  P  A  Y  G  H  H  N  V  L  N
Dt   740 TNMGNAAYACNTTYTAYGAYACNTTYTAYATHCCNGCNTAYGGNCAYCAYAAVGTNYTNA 799
         ||:::||:|| || ||| ||  :|| | :|  || || || ||   |: ||:|:: ||| 
Qy    75 TGGATGGTGGAGTTTATGATCACTTCGTCTCCACATATGGTGACCATACAGTTTTAA 134
Qt       D  G  E  F  Y  D  H  F  L  S  P  Q  Y  G  D  H  T  V  L  N
```

```
Db           S  L  A  V  I  A  L  C  H  Y  E  E  I  D  S  S  I  I  K  H
Dt   800 AYWSNYTNGCNGTNATHGCNYTNTGYCAYTAYGARGARATHGAYWSNWSNATHATHAARC 859
         : : ||:||  || || ||  ::   || || ::|| | :| ||: :||:|| || |:||
Qy   135 ATGCATTAGCTGTAATTGCACTTAAATTAGGCATGAAAAATGCCGTT 194
Qt       A  L  A  V  I  A  L  S  Y  L  E  K  L  D  V  T  N  I  K  E
```

```
Db           A  L  K  S  F  G  G  V  K  R  R  F  N  E  K  Q  L  G  D  Q
Dt   860 AYGCNYTNAARWSNTTYGGNGNGTNAARMGNMGNTTYAAYGARAAARCARYTNGGNGAYC 919
         | |:  || : :|:|| ||:||: |||:|: ::||| || :| ::| || :|:|||||:
Qy   195 AAGCATTAGAAACGTTTGGTGGTGTTAAAAGTACTTCAATGAAACTACAATTGCAAATC 254
Qt       A  L  E  T  F  G  G  V  K  R  R  F  N  E  I  T  Y  T  A  N  Q
```

```
Db           V  L  I  D  D  Y  A  H  P  T  E  I  K  V  T  I  E  A  A
Dt   920 ARGTNYTNATHGAYGAYTAYGCNCAYCCNACNGARATHAARGTNACNATHGARGCNG 979
```

```
Qy     2 TTTAAAGATTCGGATGACATTTATGCTCAAATATTCAAATTACGGATAAAGGTACTGCT 61
         |:||:||| ||:|||:||||:||||::||:||:|| |:||||:||  ||||||||| ||
Qt       F  K  D  S  D  D  I  Y  A  Q  F  Q  I  T  D  K  G  T  A
```

```
Db           F  D  V  Y  V  D  G  E  F  Y  D  H  F  L  S  P  Q  Y  G  D
Dt   727 TTYGAYGTNTAYGGNGAYGGNGARTTYTAYGAYCAYTTYYTNWSNCCNCARTAYGGNGAY 786
         |||||:||:||:||:||:||:|| || |||||:|| || ||:||:||:||:||:||:||
Qy    62 GTTGATGTGTATGGCGATGGTGAGTTTTATGATCACTTCCTGTCTCCACAAATATGGTGAC 121
Qt       V  D  V  Y  V  D  G  E  F  Y  D  H  F  L  S  P  Q  Y  G  D
```

```
Db           H  T  V  L  N  A  L  A  V  I  A  I  S  Y  L  E  K  L  D  V
Dt   787 CAYACNGTNYTNAAYGCNYTNGCNGTNATHGCNATHWSNTAYYTNGARAARYTNGAYGTN 846
         || ||:||:||:||:|| ||:|| ||:|||:||| |:||: |||||:|:||:||:||:|
Qy   122 CATACAGTTTTAAATGCATTAGCTGTAATTGCGATTAGTTATTTAGAAGCGTAGATGTT 181
Qt       H  T  V  L  N  A  L  A  V  I  A  I  S  Y  L  E  K  L  D  V
```

```
Db           T  N  I  K  E  A  L  E  T  F  G  G  V  K  R  R  F  N  E  T
Dt   847 ACNAAYATHAARGARGCNYTNGARACNTTYGGNGGNGTNAARMGNMGNTTYAAYGARACN 906
         || ||:||:||:||:|| ||:|| || ||:||:||:||:||:|: |: ||:||:|| |:
Qy   182 ACAAATATTAAAGAAGCATTAGAAACGTTTGGTGGTGTTAAACGTCGTTTCAATGAAACT 241
Qt       T  N  I  K  E  A  L  E  T  F  G  G  V  K  R  R  F  N  E  T
```

```
Db           T  I  A  N  Q  V  I  V  D  D  Y  A  H  H  P  R  E  I  S  A
Dt   907 ACNATHGCNAAYCARGTNATHGTNGAYGAYTAYGCNCAYCAYCCNMGNGARATHWSNGCN 966
         || ||:|| ||:||:||:||:||:||:||:||:|| ||:|| ||: |:||:|| |:||
Qy   242 ACAATTGCAAATCAAGTTATTGTAGATGATTATGCACACCATCCAAGAGAAATTAGTGCT 301
Qt       T  I  A  N  Q  V  I  V  D  D  Y  A  H  H  P  R  E  I  S  A
```

```
Db           T  I  E  T  A  R  K  K  Y  P  H  K  E  V  V  A  V  F  Q  P
Dt   967 ACNATHGARACNGCNMGNAARAARTAYCCNCAYAARGARGTNGTNGCNGTNTTYCARCCN 1026
         || |||||:|| || |||||||:||:||:|| ||:||:||:||:|| ||:|| ||:||:
Qy   302 ACAATTGAGACAGCACGCAAGAAAATATCCACATAAAGAAGTTGTTGCAGTATTTCAACCA 361
Qt       T  I  D  T  A  R  K  K  Y  P  H  K  E  V  V  A  V  F  Q  P
```

```
Db           H  T  F  S  R  T  Q  A  F  L  N  E  F  A  E  S  L  S  K  A
Dt  1027 CAYACNTTYWSNMGNACNCARGCNTTYYTNAAYGARTTYGCNGARWSNYTNAARAARGCN 1086
         || || ||: |:||:|| ||:|| ||: |:||:||: ||:||:| ||:||:||:||:|| 
Qy   362 CACACTTTCTCTAGAACACAAGCATTTTTAAATGAATTTGCAGAAAGTTTAAGTAAAGCA 421
Qt       H  T  F  S  R  T  Q  A  F  L  N  E  F  A  E  S  L  S  K  A
```

```
Db           D  R  V  F  L  C  E  I  F  G  S  I  R  E  N  T  G  A  L  T
Dt  1087 GAYMGNGTNTTYTNTGYGARATHTTYGGNWSNATHMGNGARAAYACNGGNGCNYTNACN 1146
         || |||||:|| | ||:|| ||:||:||:|:||:|| ||:||:|| ||:|| ||: || 
Qy   422 GATCGTGTATTCTTATGTGAAATTTTTGGATCAATTAGAGAAAATACTGGCGCCATTAACG 481
Qt       D  R  V  F  L  C  E  I  F  G  S  I  R  E  N  T  G  A  L  T
```

```
Db           I  Q  D  L  I  D  K  I  I  E  G  A  S  L  I  N  E  D  S  I  N
Dt  1147 ATHCARGAYYTNATHGAYAARATHGARGGNGCNWSNYTNATHAAYGARGAYWSNATHAAY 1206
         || ||:|| |:||:||:||:||:||:||:|| |:||:||:||:||:||: ||:|||||:
Qy   482 ATACAAGATTTAATTGATAAAATTGAAGGTGCATCGCTAATTAATGAAGATTCTATTAAT 541
Qt       I  Q  D  L  I  D  K  I  I  E  G  A  S  L  I  N  E  D  S  I  N
```

```
Db           V  L  E  Q  F  D  N  A  V  L  F  K  G  A  G  D  I  Q  K
Dt  1207 GTNYTNGARCARTTYGAYAAYGCNGTNYTNTTYAARGGNGCNGGNGAYATHCARAAR 1266
         || || |:||: || || || || ||: |||||:||:|| ||:|| |:||:|||||:
Qy   542 GTATTAGAACAATTTGATAATGCTGTGTTTAAAGGTGCAGGTGATATTCAAAAA 601
Qt       V  L  E  Q  F  D  N  A  V  L  F  M  G  A  G  D  I  Q  K
```

```
Db           L  Q  N  A  Y  L  L  G  M  K  N  A  F
Dt  1267 YTNCARAAYGCNTAYYTNGAYAARYTNGGNATGAARAAYGCNTTY 1311
         ||:||:|| || ||: |||||:||: ||:|| |||||: || || 
Qy   602 TTACAAAATGCATATTTAGATAAATTAGGCATGAAAAATGCCGTT 646
Qt       L  Q  N  A  Y  L  L  D  K  L  G  M  K  N  A  F
```

```
RESULT 2
ID   MURC_BACSU     STANDARD;     PRT;     432 AA.
AC   P40778;
DT   01-FEB-1995 (REL. 31, CREATED)
DT   15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   UDP-N-ACETYLMURAMATE--ALANINE LIGASE (EC 6.3.2.8) (UDP-N-
DE   ACETYLMURANOYL-L-ALANINE SYNTHETASE).
```

Left column

```
Qy  255 AGTTATTGTAGATGATTATGCACCATCCAAGAGAAATTAGTGCTACAATTGACACAG 314
            |:||   |:||:||:|| |||||:|| ||:||   | ||:||:||:|
Qt      V  I  V  D  D  Y  A  H  H  P  R  E  I  S  A  T  I  D  T  A

Db  980 CNMGNCARAARTAYCCNGAYMGNGARATHGTNGCNGTNTTYCARCCNCAYACNTTYACNM 1039
Dt      R  Q  K  Y  P  D  R  E  I  V  A  V  F  Q  P  H  T  F  T  R
            |:|||  |:||:|||| ||  ||  |:|| || |||| || ||
Qy  315 CACGAAGAAATATCCACATAAAGAAGTAGTTGTTGCAGATTTCAACCACACTTTCTCTA 374
Qt      R  K  K  Y  P  H  K  E  V  V  A  V  F  Q  P  H  T  F  S  R

Db 1040 GNACNCARCARTTYTNGAYGARTTYGCNGARWSNYTNWSNGGNGCNGAYTGYGTNTAYY 1099
Dt      T  Q  F  L  D  E  F  A  E  S  L  S  G  A  D  C  V  Y  L
            |  ||:|   || :||:|| |: :| |:||:|| |  |:|| |:|  |:|
Qy  375 GAACACAAGCATTTTAAATGAATTTGCAGAAGTTAAGTAAAGCAGATCGTGTATTCT 434
Qt      T  Q  A  F  L  N  E  F  A  E  S  L  S  K  A  D  R  V  F  L

Db 1100 TNTGYGAYATHTYGGNWSNGCNMGNGARAAYGCNGGNAARYTNACNATHGGNGAYYTNC 1159
Dt      C  D  I  F  G  S  A  R  E  N  A  G  K  L  T  I  G  D  L  Q
            |  |:|||| :| |||| ::|| ||:|| |:||:|| || :|
Qy  435 TATGTGAAATTTTTGGATCAATTAGAGAAAATACTGGCGCATTCAACAGAGATTTAA 494
Qt      C  E  I  F  G  S  I  R  E  N  T  G  A  L  T  I  Q  D  L  I

Db 1160 ARGGNAARATHCAYAAYGCNAARYTNATHGARGARGAYGAYACNWSNGTNYTNAARGCNC 1219
Dt      G  K  I  H  N  A  K  L  I  E  E  D  T  S  V  L  K  A  H
            |:|| |  ::||  |:|| ||:|||| :|| :| | |:|||  :|||   |
Qy  495 TTGATAAAATTGAAGGTGCATCGTAATTAATGAAGATTCTATTAATGTATTAGAACAAT 554
Qt      D  K  I  E  G  A  S  L  I  N  E  D  S  I  N  V  L  E  Q  F

Db 1220 AYGAYAARGCNCGNYTNATHTTYATGGGNGCNGGNGAYATHCARAART 1267
Dt      D  K  A  V  L  I  F  M  G  A  G  D  I  Q  K
            :|:|| || |: | || ||||||:||:|| |:|| ||:|||:|
Qy  555 TTGATAATGCTGTTGTTTATTTATGGGTGCAGGTGATATTCAAAAT 602
Qt      D  N  A  V  V  L  F  M  G  A  G  D  I  Q  K
```

Right column

```
Query Match                    9.4%;  Score 310; DB 1; Length 433;
Best Local Similarity  40.5%;  Pred. No. 1.10e-24;
Matches  77;  Conservative  38;  Mismatches  75;  Indels  0;  Gaps  0;

Qy  973 GTNYTNATHGAYGAYTAYGCNCAYCAYCCNGTNGARYTNGAYGCNGCNATHMGNWSNGTN 1032
Qt      V  L  I  D  D  Y  A  H  H  P  V  E  L  D  A  A  I  R  S  V
            :| :||:|:||:|| ||| |  || ||:|:|||:| :|:|  |:||   :|  |:||
Db  257 GTTATTGTAGATGATTATGCACACCATCCAAGAGAAATTAGTGCTACAATTGACACAGCA 316
Dt      V  I  V  D  D  Y  A  H  H  P  R  E  I  S  A  T  I  D  T  A

Qy 1033 MGNGARATHTAYWSNGGNAARCAYATHATGGGNATHTTYCARCCNCAYYTNTAYWSNMGN 1092
Qt      R  E  I  Y  S  G  K  H  I  M  G  I  F  Q  P  H  L  Y  S  R
            :|:||  |:| |  |:||  || ||::| |  |:||  ||:| | ||:| || |||
Db  317 CGAAAGAAATATCCACATAAAGAAGTAGTTGTTGCAGTATTCAACCACACTTTCTCTAGA 376
Dt      R  K  K  Y  P  H  K  E  V  V  A  V  F  Q  P  H  T  F  S  R

Qy 1093 ACNGCNGAYTTYTAYCARGAYTTYGCNAARWSNYTNWSNATGYTNGAYCARGTNGTNYTN 1152
Qt      T  A  D  F  Y  Q  D  F  A  K  S  L  S  M  L  D  Q  V  V  L
            ||:|||:| ||:||:|| || ||||:| ::|||:|:|  | || ||:||:||:||
Db  377 ACACAAGCATTTTAAATGAATTTGCAGAAGTTAAGTTAAGTAAAAGCAGATCGTGTATTCTTA 436
Dt      T  Q  A  F  L  N  E  F  A  E  S  L  S  K  A  D  R  V  F  L

Qy 1153 YTNGAYATHT 1162
Qt      L  D  I
            |::|:|
Db  437 TGTGAAATTT 446
Dt      C  E  I
```

```
Db  973 GTNYTNATHGAYGAYTAYGCNCAYCAYCCNGTNGARYTNGAYGCNGCNATHMGNWSNGTN 1032
Db 1033 MGNGARATHTAYWSNGGNAARCAYATHATGGGNATHTTYCARCCNCAYYTNTAYWSNMGN 1092
Db 1093 ACNGCNGAYTTYTAYCARGAYTTYGCNAARWSNYTNWSNATGYTNGAYCARGTNGTNYTN 1152
Db 1153 YTNGAYATHT 1162
```

```
                    /protein_id="AAB87090.1"
                    /db_xref="PID:g2642659"
                    /db_xref="GI:2642659"
                    /translation="MTHYHFVGIKGSGMSSLAQIMHDLGHEVQGSDIENVFTEVALR
                    NKGIKILPDANNIKEDMVVIQGNAFASSHEEKARAHQMKLDVVSYNDFLGQIIDQYT
                    SVAVTGAHGKTSTTGLLSHVMNGDKKTSFLIGDGVGMGLPESDYFAFEACEYRRHFLS
                    YKPDYAIMTNIDFDHPDYFKDINDVFDAFQEMAHNVKKGIIAWGDEHLRKIEADVPI
                    YYYGFKDSDDIYAQNIQITDKGTAFDVYVDGEFYDHFLSPQYGDHTVLNALAVIAISY
                    LEKLDVTNIKEALETFGGVKRRFNETTIANQVIVDDYAHHPREISATIETARKKIYHK
                    EVVAVFQPHTFSRTQAFLNEFAESLSKADRVFLCEIFGSIRENTGALTIQDLIDKIEG
                    ASLINEDSINVLEQFDNAVVLFKGAGDIQKLQNAYLDKLGMKNAF"

BASE COUNT     467 a      163 c      250 g      434 t
ORIGIN

Query Match    96.6%;    Score 2573;    DB 17;    Length 1314;
Best Local Similarity  59.5%;    Pred. No. 8.76e-275;
Matches  384;  Conservative  158;  Mismatches  103;  Indels  0;  Gaps  0;

Db     667  TTTAAAGATTCCGATGACATTTATGCTCAAAAATATTCAAATTACGGATAAAGGTACTGCT  726
             || | ||| : || :|| || :|||| || :  |:||:|||:| ||:|| :||  ||
Qy       1  TTYAARGAYWSNGAYGAYATHTAYGCNCARATHTYCARATHACNGAYAARGGNACNGCN    60

Db     727  TTTGATGTGTATGTCGGATGGTGAGTTTATGATCACTTCCTGTCTCCACAATATGGTGAC  786
             || || ||||| || || | ||  :|| :||  ||  |||| ||:||  ||  |:
Qy      61  GTNGAYGTNTAYGTNGGNGARTTYTAYGAYCAYTTYYTNWSNCCNCARTAYGGNGAY   120

Db     787  CATACAGTTTAAATGCATTAGCTGTAATTGCGATTAGTTATTTAGAGAAGCTAGATGTT   846
             ||:|| ||  |||:|| || | |  || | :| || ::|:  |:|| |||  |:|:||
Qy     121  CAYACNGTNTNAAYGCNYTNGCNGTNATHGCNATHWSNTAYTNGARARAYTNGAYGTN   180

Db     847  ACAAATATTAAAGAAGCATTAGAAACGTTTGGTGGTGTTAAACGTCGTTCAATGAAACT   906
             ||:|| |  |||| | || | |  |:||:||:||:|| ||  ||:  |:|||||  ||
Qy     181  ACNAAYATHAARGARGCNYTNGARACNTTYGGNGGNGTNAARMGNMGNTTYAAYGARACN  240

Db     907  ACAATTGCAAATCAAGTTATTGTAGATGATTGATGATGCACACCATCCAAGAGAAAGTTAGTGCT  966
             || |||||||| |||| |||| || ||:|| :| || ||  ||::||:||:|| :|:|:||
Qy     241  ACNATHGCNAAYCARGTNATHGTNGAYGAYATHGCNCAYCCNMGNGARATHWSNGCN   300

Db     967  ACAATTGAAACAGCACGACGAAAGAAATATCCACATAAAGAAGTTGTTGCAGTATTCAACCA  1026
             || |||||||| || || || |||  ||:|| |||:|| || || |||||::|:|||| ||
Qy     301  ACNATHGAYACNGCNMGNACNGCNAARAARTAYCCNCAYARRGARGTNGTNGCNGTNTTYCARCCN  360

Db    1027  CACACTTCTCTAGAACACAAGCACATTTTTAAATGAATTCGCAGAAAGTTTAAGTAAAGCA  1086
             ||:||:||:||:||:| :| |  |||  |||:||:|| ||  ||||||:||||: | |||
Qy     361  CAYACNTTYWSNMGNACNCARGCNTTYTNAAYGARTTYGCNGARWSNYTNWSNAARGCN   420

Db    1087  GATCGTGTATTCTTTATGTGAAATTTTGGATCAATTAGAGAAAATACTGGCGCCATTAACG  1146
             || || |||| || |||||  :|: || |||:| | | |  || ||:||:||:|| ||
Qy     421  GAYMGNGTNTTYTTYATNTGYGARATHTTYGGNWSNATHMGNGARAAYACNGGNGCNYTNACN  480

Db    1147  ATACAAGATTTAATTGATAAAATTGAAGGTGCATCGTTAATTAATGAAGATTCCATTAAT  1206
             || ||||| ||| || || || ||||| || || | ||:||:|| || || :| || |||
Qy     481  ATHCARGAYYTNATHGAYAARATHGARGGNGCNWSNYTNATHAAYGARGAYWSNATHAAY   540

Db    1207  GTATTAGAACAAATTGATAATGCTGTCTGTTTTATTTAAGGTGCAGGTGATATTCAAAAA  1266
             || |||||:||:||:|| || ||:|| |  |||:||::|| ||||| || || |||||
Qy     541  GTNYTNGARCARYTYGAYAAYGCNGTNGTNTTYATHWSNGGNGCNGGNGAYATHCARAAR   600

Db    1267  TTACAAAATGCATATTTAGATAAAATTAGGCCATGAAAAATGCGTTT  1311
             :| ||||| || || ||| ||:|| ||| : ||:|||||  || |
Qy     601  YTNCARAAYGCNTAYYTNGAYAARYTNGGNATGAARAAAYGCNTTY   645
```

```
REFERENCE   Eubacteria; Firmicutes; Low G+C gram-positive bacteria;
            Bacillaceae; Bacillus.
AUTHORS     1 (bases 1 to 2320)
TITLE       Varon,D., Brody,M.S. and Price,C.W.
            Bacillus subtilis operon under the dual control of the general
            stress transcription factor sigma B and the sporulation
            transcription factor sigma H
JOURNAL     Mol. Microbiol. 20 (2), 339-350 (1996)
MEDLINE     96310371
FEATURES             Location/Qualifiers
     source          1..2320
                     /organism="Bacillus subtilis"
                     /strain="168"
                     /sub_species="Marburg"
                     /db_xref="taxon:1423"
     CDS             <1..1049
                     /note="identification based on sequence similarity to E.
                     coli enzyme"
                     /codon_start=3
                     /transl_table=11
                     /product="UDP-N-acetyl muramate-alanine ligase"
                     /protein_id="AAB40043.1"
                     /db_xref="PID:g556014"
                     /db_xref="GI:556014"
                     /translation="IPVIRYHKFLGDYMKKFTSVAVTGAHGKTSTGLLAHVIQNAKP
                     TSFLIGDGTGQNENSEYFVFEACEYRRHFLSYQPDYAIMTNIDFDHPDYFSSIDDVF
                     DAFQEMALQVNKGIIACGDDEHLPKIHANVPVVYGTGEENDFQARNIVKSTEGTTFD
                     VEVRNIFYDFYIPAYGHHNVLNSLAVIALCHYEEIDSSIIKHALKSFGGVKRRFNEK
                     QLGDQVLIDDYAHHPTEIKVTIEAARQKYPDREIVAVFQPHTFRTQQFLDEFAESLS
                     GADCVYLCDIFGSARENAGKLTIGDLQGKIHNAKLIEEDDTSVLKAHDKAVLIFMGAG
                     DIQKYMRAYENVMA"
     terminator      1048..1079
     -35_signal      1101..1109
                     /note="sigma-H consensus"
     -10_signal      1123..1127
                     /note="sigma-H consensus"
     prim_transcript 1136
                     /note="sigma-H dependent 5' end"
     -35_signal      1151..1158
                     /note="sigma-B consensus"
     -10_signal      1172..1177
                     /note="sigma-B consensus"
     prim_transcript 1186
                     /note="sigma-B dependent 5' end"
     RBS             1198..1205
     CDS             1211..1642
                     /note="ORF1"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAB40044.1"
                     /db_xref="PID:g556015"
                     /db_xref="GI:556015"
                     /translation="MIIILYLSVALIAVAFLVLVIYLSKTLKSLQLTLKNVASTLEGL
                     EGQMKGITTETAELLHKTNRLAEDIQEKSEKLNTVHAVQGVGASVQQFNTSMKQAAG
                     SVSASVRENQDKINQVVGRSTWSQAAMEIWEKWKQKKKSAL"
     RBS             1660..1665
     CDS             1670..2128
                     /note="similar to plant water stress proteins; ORF2"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAB40045.1"
                     /db_xref="PID:g556016"
                     /db_xref="GI:556016"
                     /translation="MMSKDGINSKDFLIGTLIGGIIGATTALFLAPKSGKELRDDLGS
                     QAVALRDKTDKMTADAKEKGTQYVSIAKDKTSNIIQLVADQSQIMNKVKDLRDRSKS
                     DKTDSSTAMQDMREEAMQAADETKDQVLQTKEDVKEDVKDLKDAQKQAEQLNR"
     RBS             2135..2139
     CDS             2152..>2320
                     /note="open reading frame extends an additional 53 codons
                     (including the TGA stop) into the overlapping sequence
                     reported by Bolotin et"
                     /codon_start=1
                     /transl_table=11
```

```
RESULT       2
LOCUS        BACUNAM      2320 bp     DNA           BCT       26-MAR-1997
DEFINITION   Bacillus subtilis UDP-N-acetylmuramate-alanine ligase gene, partial
             cds, and 3 ORF's.
ACCESSION    L31845
NID          g556013
VERSION      L31845.1   GI:556013
KEYWORDS     UDP-N-acetylmuramate:L-alanine ligase.
SOURCE       Bacillus subtilis
ORGANISM     Bacillus subtilis
```

```
/protein_id="AAB40046.1"
/db_xref="PID:g556017"
/db_xref="GI:556017"
/translation="MAKQLIQSEEFKRIAEQEGVFVFLKHSTTCPISQAAFHEFDAF
ANQHEDVPAYYL"
```

```
BASE COUNT      750 a     464 c     549 g     557 t
ORIGIN

Query Match         49.2%;    Score 1309;   DB 17;    Length 2320;
Best Local Similarity  45.2%;  Pred. No. 3.43e-125;
Matches  272;  Conservative  123;  Mismatches 207;  Indels  0;  Gaps  0;

Db   430 ATGATTTCAAGCCAGCCAGAAACATCGTAAAAAGCACGGAGGGACAACTTTTGATGTCTTTG 489
         |:||:| ||  :||   :|:: :| |  ||  || |   || ::|||||||:|
Qy    14 AYGAYATHTAYGCNCARATHTTYCARATHACNGAYAARGNACNCGNTNGAYGTNTAYG 73

Db   490 TCCGCAATACGTTCTATGATACGTTTTATATTCCTGCGTACGGCCACCACAATGTATTAA 549
         :||:|||:|| ||:: ||   :|| :|||:|:|||    |::|||| || |
Qy    74 TNGAYGGNGARTTYATYGAYCAYTTYTNWSNCCNCARTAYGGNGAYCAYACNGTNYTNA 133

Db   550 ACTCATTGGCGGTCATTGCGTTATGCCATTATGAAGAAATTGATTCCAGCATCATTAAGC 609
         |: :: || |||| ||:|| ||:|| |||| || |    || :||||   ||| |
Qy   134 AYGCNYTNGCNGTNATHGCNATHWSNTAYYTNGARAARYTNGAYGTNACNAAYATHAARG 193

Db   610 ATGCTCTCAAATCCTTTGGGGGCGTCAAACGACCAGATTCAATGAGAAGCAGCTTGGGGATC 669
         | ||  :||  |::||| || || || :: |  :|| |:| :|:|||   ::|
Qy   194 ARGCNYTNGARACNTTYGGNGNGTNAARMGNMGNTTYAAYGARACNACNATHGCNAAYC 253

Db   670 AAGTGCTGATTGATGACTATGCCCATCATCCGCACAGAAATAAAAGTGACAAATCGAGGCGG 729
         |:|| |||| |:|:|::|| ||:||:| : |||||: | ||:| |
Qy   254 ARGTNATHGTNGAYGAYTAYGCNCAYCAYCCNMGNGARATHWSNGCNACNATHGAYACNG 313

Db   730 CAAGACAGAAGTATCCTGATCGGGAAATTGTCGCGGTATTCCAGCCTCATACATTTACCC 789
         | : |:||:|||| || :: :|| ::||:||| || :: :|| |||:|||  |
Qy   314 CNMGNAARAARTAYCCNCAYAARGARGTNGTNGCNGTNTTYCARCCNCAYACNTTYWSNM 373

Db   790 GGACGCAGCAGTTCCTTGACGAATTGCAGAAAGCCTGAGCGGCGCCGACTGTGTGTATT 849
         | | ||:   ||||:||:|| | ||:|:| ||   |:|  ||| |  || :|
Qy   374 GNACNCARGCNTTYTNAAYGARTTYGCNGARWSNYTNWSNAARGCNGAYMGNGTNTTYY 433

Db   850 TATGCGATATTTCGGCTCAGCCCGTGAGAATGCCGGAAAGCTGACGATCGGTGATCTCC 909
         | |:|| ||:|| ||:::  : ::||||| ||| |   |:|||  |:::|
Qy   434 TNTGYGARATHTYGGNWSNATHMGNGARAAYACNGGNGCNYTNACNATHCARGAYYTNA 493

Db   910 AGGGAAAAATTCATAATGCCAAGCTGATTGAAGAAGATGACACATCGTTTTAAAAGCTC 969
         | ||:| ||:| :: || ::: :|:: :|| | :|| :||:|
Qy   494 THGAYAARATHGARGGNGCNWSNYTNATHAAYGARGAYWSNATHAAYGTNYTNGARCART 553

Db   970 ATGATAAAGCCCGTTCCTCATCTTCATGGGAGCAGGAGATATCCAAAAAATATGAGAGCCT 1029
         :||:|| || || :||| ||||  |||| ||| |   |:: | || :
Qy   554 TYGAYAAYGCNGTNGTNYTNTTYATGGGNGCNGGNGAYATHCARAAARYTNCARAAYGCNT 613

Db  1030 AC 1031
         |:
Qy   614 AY 615
```

```
RESULT     3
LOCUS       BSUB0016    215640 bp    DNA                   BCT       26-NOV-1997
DEFINITION  Bacillus subtilis complete genome (section 16 of 21); from 2997771
            to 3213410.
ACCESSION   Z99119 AL009126
NID         g2635411
VERSION     Z99119.1  GI:2635411
KEYWORDS    .
SOURCE      Bacillus subtilis.
ORGANISM    Bacillus subtilis
            Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillaceae;
            Bacillus.
REFERENCE   1 (bases 1 to 215640)
AUTHORS     Kunst,F., Ogasawara,N., Moszer,I., Albertini,A.M., Alloni,G.,
            Azevedo,V., Bertero,M.G., Bessieres,P., Bolotin,A., Borchert,S.,
            Borriss,R., Boursier,L., Brans,A., Braun,M., Brignell,S.C.,
            Bron,S., Brouillet,S., Bruschi,C.V., Caldwell,B., Capuano,V.,
            Carter,N.M., Choi,S.K., Codani,J.J., Connerton,I.F., Cummings,N.J.,
            Daniel,R.A., Denizot,F., Devine,K.M., Dusterhoft,A., Ehrlich,S.D.,
            Emmerson,P.T., Entian,K.D., Errington,J., Fabret,C., Ferrari,E.,
            Foulger,D., Fritz,C., Fujita,M., Fujita,Y., Fuma,S., Galizzi,A.,
            Galleron,N., Ghim,S.Y., Glaser,P., Goffeau,A., Golightly,E.J.,
            Grandi,G., Guiseppi,G., Guy,B.J., Haga,K., Haiech,J., Harwood,C.R.,
            Henaut,A., Hilbert,H., Holsappel,S., Hosono,S., Hullo,M.F.,
            Itaya,M., Jones,L., Joris,B., Karamata,D., Kasahara,Y.,
            Klaerr-Blanchard,M., Klein,C., Kobayashi,Y., Koetter,P.,
            Koningstein,G., Krogh,S., Kumano,M., Kurita,K., Lapidus,A.,
            Lardinois,S., Lauber,J., Lazarevic,V., Lee,S.M., Levine,A., Liu,H.,
            Masuda,S., Mauel,C., Medigue,C., Medina,N., Mellado,R.P.,
            Mizuno,M., Moestl,D., Nakai,S., Noback,M., Noone,D., O'Reilly,M.,
            Ogawa,K., Ogiwara,A., Oudega,B., Park,S.H., Parro,V., Pohl,T.M.,
            Portetelle,D., Porwollik,S., Prescott,A.M., Presecan,E., Pujic,P.,
            Purnelle,B., Rapoport,G., Rey,M., Reynolds,S., Rieger,M.,
            Rivolta,C., Rocha,E., Roche,B., Rose,M., Sadaie,Y., Sato,T.,
            Scanlan,E., Schleich,S., Schroeter,R., Scoffone,F., Sekiguchi,J.,
            Sekowska,A., Seror,S.J., Serror,P., Shin,B.S., Soldo,B.,
            Sorokin,A., Tacconi,E., Takagi,T., Takahashi,H., Takemaru,K.,
            Takeuchi,M., Tamakoshi,A., Tanaka,T., Terpstra,P., Tognoni,A.,
            Tosato,V., Uchiyama,S., Vandenbol,M., Vannier,F., Vassarotti,A.,
            Viari,A., Wambutt,R., Wedler,E., Wedler,H., Weitzenegger,T.,
            Winters,P., Wipat,A., Yamamoto,H., Yamane,K., Yasumoto,K., Yata,K.,
            Yoshida,K., Yoshikawa,H.F., Zumstein,E., Yoshikawa,H. and
            Danchin,A.
TITLE       The complete genome sequence of the gram-positive bacterium
            Bacillus subtilis
JOURNAL     Nature 390 (6657), 249-256 (1997)
MEDLINE     98044033
REFERENCE   2 (bases 1 to 215640)
AUTHORS     Kunst,F., Ogasawara,N., Yoshikawa,H. and Danchin,A.
TITLE       Direct Submission
JOURNAL     Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur,
            Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724
            Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr,
            adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45 45
            68 89 48
FEATURES             Location/Qualifiers
     source          1..215640
                     /organism="Bacillus subtilis"
                     /strain="168"
                     /db_xref="taxon:1423"
     terminator      complement(8..24)
                     /gene="ytnM"
     gene            complement(8..1001)
                     /gene="ytnM"
     CDS             complement(99..1001)
                     /gene="ytnM"
                     /function="unknown"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAB14906.1"
                     /db_xref="PID:e1185801"
                     /db_xref="PID:g2635412"
                     /db_xref="GI:2635412"
                     /db_xref="SPTREMBL:O34430"
                     /translation="MKKLIVFAFIGLLSOLIDGSLGMAYGVTSTSLLLAFGITPAVAS
                     ASVHLAEVVTTAASGVSHIKFGNVDKQTVYQLVIPGSIGAFLGAAFLSQLPGDVAKPY
                     ISLFLLLGGYVLRFLFQYKPALEKKHVPLNRKQSIPLGVIAGFADATGGGWGPVT
                     TPILLSRKGLSPRKVVGTVDTSEPAIAVSATAGFLISLGEDVNWLWVFSLMAGGIIA
                     APIAAWLVQKFHPQLMGVLVGGFIILVNARTLINEWIANTAVHPLIYTAIGAIWLSAV
                     LFVLSKIGNRNIVKTAVDVHLKEK"
     gene            complement(1020..2270)
                     /gene="hipO"
     CDS             complement(1020..2270)
                     /gene="hipO"
                     /EC_number="3.5.1.32"
                     /codon_start=1
                     /transl_table=11
                     /product="hippurate hydrolase"
                     /protein_id="CAB14907.1"
```

Qy 61 TNIKEALETFGGVKRRFNETTIANQVIVDDYAHHPREISATIDTARKKYPHKEVVAVFQP 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 343 HTFSRTQAFLNEFAESLSKADRVFLCEIFGSIRENTGALTIQDLIDKIEGASLINEDSIN 402

Qy 121 HTFSRTQAFLNEFAESLSKADRVFLCEIFGSIRENTGALTIQDLIDKIEGASLINEDSIN 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db 403 VLEQFDNAVVLFKGAGDIQKLQNAYLDKLGMKNAF 437
        |||||||||||||||||||||||||||||
Qy 181 VLEQFDNAVVLFMGAGDIQKLQNAYLDKLGMKNAF 215

RESULT 2
ID   MURC_BACSU   STANDARD;   PRT;   432 AA.
AC   P40778;
DT   01-FEB-1995 (REL. 31, CREATED)
DT   15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   UDP-N-ACETYLMURAMATE--ALANINE LIGASE (EC 6.3.2.8) (UDP-N-
DE   ACETYLMURANOYL-L-ALANINE SYNTHETASE).
GN   MURC.
OS   BACILLUS SUBTILIS.
OC   BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC   BACILLUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 98048467.
RA   LAPIDUS A., GALLERON N., SOROKIN A., EHRLICH S.D.;
RT   "Sequencing and functional annotation of the Bacillus subtilis genes
RT   in the 200 kb rrnB-dnaB region.";
RL   MICROBIOLOGY 143:3431-3441(1997).
RN   [2]
RP   SEQUENCE OF 85-432 FROM N.A.
RC   STRAIN=168 / MARBURG;
RX   MEDLINE; 96310371.
RA   VARON D., BRODY M.S., PRICE C.W.;
RT   "Bacillus subtilis operon under the dual control of the general
RT   stress transcription factor sigma B and the sporulation transcription
RT   factor sigma H.";
RL   MOL. MICROBIOL. 20:339-350(1996).
-!- FUNCTION: CELL WALL FORMATION.
-!- CATALYTIC ACTIVITY: ATP + UDP-N-ACETYLMURAMOYL-L-ALANINE -
    ADP + ORTHOPHOSPHATE + UDP-N-ACETYLMURAMOYL + L-ALANINE.
-!- PATHWAY: PEPTIDOGLYCAN BIOSYNTHESIS.
-!- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).
-!- SIMILARITY: BELONGS TO THE MURCDEF FAMILY.

DR   EMBL; AF008220; G2293216; -.
DR   EMBL; L31845; G556014; -.
DR   SUBTILIST; BG10973; MURC.
KW   PEPTIDOGLYCAN SYNTHESIS; CELL WALL; CELL DIVISION; LIGASE;
KW   ATP-BINDING.
FT   NP_BIND   108   114   ATP (POTENTIAL).
SQ   SEQUENCE   432 AA; 48364 MW; 163166CB CRC32;

Query Match
Best Local Similarity   57.7%;   Score: 863;   DB 1;   Length 432;
Matches 120;   Conservative   32;   Mismatches 51;   Indels 0;   Gaps 0;

Db 225 EENDFQARNIVKSTEGTTFDVFVRNTFYDTFYIPAYGHHNVLNSLAVIALCHYEEIDSSI 284
Qy   3 DSDDIYAQIFQITDKGTAVDVVVDGEFYDHFLSPQYGDHTVLNALAVIAISYLEKLDVTN 62

Db 285 IKHALKSFGGVKRRFNEKQLGDQVLIDDYAHHPTEIKVTIEAARQKYPDREIVAVFQPHT 344
Qy 63 IKEALETFGGVKRRFNETTIANQVIVDDYAHHPREISATIDTARKKYPHKEVVAVFQPHT 122

Db 345 FTRTQQFLDEFAESLSGADCVYLCDIFGSARENAGKLTIGDLQGKIHNAKLIEEDDTSVL 404
Qy 123 FSRTQAFLNEFAESLSKADRVFLCEIFGSIRENTGALTIQDLIDKIEGASLINEDSINVL 182

Db 405 KAHDKAVLIFMGAGDIQKYMRAY 427
Qy 183 EQFDNAVVLFMGAGDIQKLQNAY 205

RESULT 3
ID   MURC_SYNY3   STANDARD;   PRT;   505 AA.
AC   P74528;
DT   15-JUL-1998 (REL. 36, CREATED)
DT   15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   UDP-N-ACETYLMURAMATE--ALANINE LIGASE (EC 6.3.2.8) (UDP-N-
DE   ACETYLMURANOYL-L-ALANINE SYNTHETASE).
GN   MURC OR SLR1423.
OS   SYNECHOCYSTIS SP. (STRAIN PCC 6803).
OC   BACTERIA; CYANOBACTERIA; CHROOCOCCALES; SYNECHOCYSTIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97061201.
RA   KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
RA   MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,
RA   HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARUO K., OKUMURA S.,
RA   SHIMPO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M.,
RA   TABATA S.;
RT   "Sequence analysis of the genome of the unicellular cyanobacterium
RT   Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT   entire genome and assignment of potential protein-coding regions.";
RL   DNA RES. 3:109-136(1996).
-!- FUNCTION: CELL WALL FORMATION (BY SIMILARITY).
-!- CATALYTIC ACTIVITY: ATP + UDP-N-ACETYLMURAMOYL + L-ALANINE -
    ADP + ORTHOPHOSPHATE + UDP-N-ACETYLMURAMOYL-L-ALANINE.
-!- PATHWAY: PEPTIDOGLYCAN BIOSYNTHESIS.
-!- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).
-!- SIMILARITY: BELONGS TO THE MURCDEF FAMILY.

DR   EMBL; D90916; G1653723; -.
KW   PEPTIDOGLYCAN SYNTHESIS; CELL WALL; CELL DIVISION; LIGASE;
KW   ATP-BINDING.
FT   NP_BIND   164   170   ATP (POTENTIAL).
SQ   SEQUENCE   505 AA; 54513 MW; B3D7FC7C CRC32;

Query Match
Best Local Similarity   22.9%;   Score: 343;   DB 1;   Length 505;
Matches 57;   Conservative   29;   Mismatches 40;   Indels 4;   Gaps 4;

Db 322 GDHNISNALAAVAVGRLLGLDFPVIAQAIASFNGAKRRFECKGCNGITFIDDYAHHPSE 381
Qy 39 GDHTVLNALAVIAISYLEKLDVTNIKEALETFGGVKRRFNETTIANQV-IVDDYAHHPRE 97

Db 382 LLATLAAAKQKVTHGKYERVVAIFQPHRYSRTHTFMAEFATAFKDADLVLTDIYSAGEQ 441
Qy 98 ISATIDTARKKYPH-K-E-VVAVFQPHTFSRTQAFLNEFAESLSKADRVFLCEIFGSIRE 154

Db 442 NPYNIRGEDL 451
Qy 155 NTGALTIQDL 164

RESULT 4
ID   MURC_PORGI   STANDARD;   PRT;   433 AA.